

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 16:38:49 ; Search time 56 Seconds
(without alignments)
4432.963 Million cell updates/sec

Title: US-09-734-672-4
Perfect score: 9649
Sequence: 1 MDLSALRVEYQNVINAMQK.....LYQCQLDYLIPQIPHSHY 1863

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9649	100.0	1863	17 AAR81535	BRCA1 mutant from
2	9649	100.0	1863	19 AAW76099	Human BRCA1 (Omi2 p
3	9649	100.0	1863	21 AAB24219	Human BRCA1 (Omi3)
4	9642	99.9	1863	17 AAR97128	BRCA1, breast and
5	9642	99.9	1863	17 AAR81481	BRCA1, Homo sapie
6	9642	99.9	1863	17 AAR91208	BRCA1, breast and
7	9642	99.9	1863	20 AAY32033	Human BRCA1 protei
8	9642	99.9	2353	22 ABG01696	Novel human digno
9	9639	99.9	1863	17 AAR81490	BRCA1 mutant from
10	9639	99.9	1863	17 AAR81500	BRCA1 mutant from

11	9639	99.9	1863	17 AAR81522	BRCA1 mutant from
12	9639	99.9	1863	17 AAR81536	BRCA1 mutant from
13	9639	99.9	1863	17 AAR81540	BRCA1 mutant from
14	9638	99.9	1863	17 AAR81529	BRCA1 mutant from
15	9638	99.9	1863	17 AAR81532	BRCA1 mutant from
16	9638	99.9	1863	17 AAR81533	BRCA1 mutant from
17	9638	99.9	1863	17 AAR81534	BRCA1 mutant from
18	9638	99.9	1863	17 AAR81514	BRCA1 mutant from
19	9638	99.9	1863	17 AAR81543	BRCA1 mutant from
20	9638	99.9	1863	17 AAR81545	BRCA1 mutant from
21	9637	99.9	1863	17 AAR81485	BRCA1 mutant from
22	9637	99.9	1863	17 AAR81493	BRCA1 mutant from
23	9637	99.9	1863	17 AAR81505	BRCA1 mutant from
24	9637	99.9	1863	17 AAR81511	BRCA1 mutant from
25	9637	99.9	1863	17 AAR81538	BRCA1 mutant from
26	9636	99.9	1863	17 AAR81486	BRCA1 mutant from
27	9636	99.9	1863	17 AAR81509	BRCA1 mutant from
28	9636	99.9	1863	17 AAR81537	BRCA1 mutant from
29	9636	99.9	1863	17 AAR81542	BRCA1 mutant from
30	9636	99.9	1863	17 AAR81544	BRCA1 mutant from
31	9635	99.9	1863	17 AAR81524	BRCA1 mutant from
32	9635	99.9	1863	17 AAR81546	BRCA1 mutant from
33	9635	99.9	1863	18 AAW25522	Human BRCA1 consen
34	9635	99.9	1863	19 AAW79665	BRCA1 (Omi1) prote
35	9635	99.9	1863	19 AAW76100	Human BRCA1 (Omi3 p
36	9635	99.9	1863	19 AAW76098	Human BRCA1 (Omi1 p
37	9635	99.9	1863	21 AAB24217	Human BRCA1 (Omi1)
38	9635	99.9	1863	21 AAB24218	Human BRCA1 (Omi2)
39	9634	99.8	1863	17 AAR81488	BRCA1 mutant from
40	9634	99.8	1863	17 AAR81519	BRCA1 mutant from
41	9634	99.8	1863	17 AAR81539	BRCA1 mutant from
42	9630	99.8	1863	17 AAR99440	BRCA1 allele #8403
43	9630	99.8	1863	17 AAR81497	BRCA1 mutant from
44	9630	99.8	1863	18 AAW10003	Protein encoded by
45	9630	99.8	1863	19 AAW79889	Tumourogenic BRCA1

ALIGNMENTS

RESULT 1
AAR81535
ID AAR81535 standard; Protein; 1863 AA.
XX AC AAR81535;
XX DT 02-OCT-1996 (first entry)
XX DE BRCA1 mutant from PM06.
XX KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Misc-difference 871 /note= "p871L"
XX PN WO9605306-A2.
XX PD 22-FEB-1996.
XX PF 11-AUG-1995; 95WO-US10202.
XX PR 07-JUN-1995; 95US-0483553.
PR 12-AUG-1994; 94US-0289221.
PR 02-SEP-1994; 94US-0300266.
PR 16-SEP-1994; 94US-0308104.
PR 29-NOV-1994; 94US-0348824.
PR 24-MAR-1995; 95US-0409305.
PR 07-JUN-1995; 95US-0480784.

Qy 1561 TPYLESGISLFSDDPESDRAPE SARVGNIPSSTSALKVPQLKVAESAQSPAAAHTT 1620
Db 1561 TPYLESGISLFSDDPESDRAPE SARVGNIPSSTSALKVPQLKVAESAQSPAAAHTT 1620
Qy 1621 DTAGYNAMESVSREKPELTASTERVNKRSMVMVSGLTPEEFMLVYKFARKHHTLTNLI 1680
Db 1621 DTAGYNAMESVSREKPELTASTERVNKRSMVMVSGLTPEEFMLVYKFARKHHTLTNLI 1680
Qy 1681 TEETHVVMKTDAEFVCERTLKYLFLGIAGGKWVYSYFWVTQSIKERKMLNEHDFEVRGDV 1740
Db 1681 TEETHVVMKTDAEFVCERTLKYLFLGIAGGKWVYSYFWVTQSIKERKMLNEHDFEVRGDV 1740
Qy 1741 VNGRHOGPKRARESQRKIFRGLIEICCYGFTNMPTDQLEWMYQLCGASVVKLSLFTL 1800
Db 1741 VNGRHOGPKRARESQRKIFRGLIEICCYGFTNMPTDQLEWMYQLCGASVVKLSLFTL 1800
Qy 1801 GTGVHPITVVVQPDATWEDNGFHAIGOMCEAPVVTREWVLDSVALYQCOELDTYLIPQIPH 1860
Db 1801 GTGVHPITVVVQPDATWEDNGFHAIGOMCEAPVVTREWVLDSVALYQCOELDTYLIPQIPH 1860
Qy 1861 SHY 1863
Db 1861 SHY 1863
RESULT 2
AAW76099
ID AAW76099 standard; Protein; 1863 AA.
XX
AC AAW76099;
XX
DT 18-NOV-1998 (first entry)
XX
DE Human BRCA1 omi2 protein.
XX
KW BRCA1; omi2; human; breast and ovarian cancer predisposing gene;
KW polymorphism; susceptibility; anti-oncogene; tumour suppressor;
XX chromosome 17q.
XX
OS Homo sapiens.
XX
PN US5750400-A.
XX
PD 12-MAY-1998.
XX
PF 12-FEB-1997; 97US-0798691.
XX
PR 12-FEB-1996; 96US-0598591.
PR 12-FEB-1997; 97US-0798691.
XX
PA (ONCO-) ONCORMED INC.
XX
PI Allen AC, Alvares CP, Critz BS, Murphy PD, Olson SJ;
PI Schelter DB, Zeng B;
XX
DR WPI; 1998-296774/26.
DR N-PSDB; AAV46449.
XX
PT BRCA1 omi gene coding sequences - useful for distinguishing between
PT polymorphisms and mutation(s) in the screening for disposition to
PT breast or ovarian cancer
XX
PS Disclosure; Column 47-56; 54pp; English.
XX
CC This sequence represents the human BRCA1 (breast and ovarian cancer
CC predisposing gene) omi2 gene. This sequence and polymorphic variations of
CC this sequence are useful for the identification of an individual who may
CC or may not have an increased susceptibility to breast or ovarian cancer.
CC The sequences used identify gene changes which are due to polymorphisms,
CC rather than cancer causing mutations. BRCA1 is an anti-oncogene (tumour
CC suppressor) which is involved in genetic inheritance of cancers,
CC especially breast and ovarian cancer. It is found at human chromosome 17q
CC which is known to be linked to cancer susceptibility, especially breast

CC cancer. Cells containing a mutation in this gene lose the wild-type
CC function of BRCA1 and are more susceptible to cancers.
XX
SQ Sequence 1863 AA;
Query Match 100.0%; Score 9649; DB 19; Length 1863;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MDLSALRVEEVQNVINAMOKILECPICLELILKEPVSSTKCDHIFCKFCMLLKNQKGPQ 60
Db 1 MDLSALRVEEVQNVINAMOKILECPICLELILKEPVSSTKCDHIFCKFCMLLKNQKGPQ 60
Qy 61 CPICKNDITKRSLQESTRFSQLYVEELKIIICAFOLDTGLEYANSYNPAKKENNSPEHLKD 120
Db 61 CPICKNDITKRSLQESTRFSQLYVEELKIIICAFOLDTGLEYANSYNPAKKENNSPEHLKD 120
Qy 121 EVSTIQSMGYRNRAKRLQLQSEPNPSLQETSLSVQLSNLGTVRTLRKQIQPKTSVYI 180
Db 121 EVSTIQSMGYRNRAKRLQLQSEPNPSLQETSLSVQLSNLGTVRTLRKQIQPKTSVYI 180
Qy 181 ELGSDSSEDTVNKATYCSVGDOELLIQITPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDTVNKATYCSVGDOELLIQITPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Qy 241 PSNNDLNTTEKRAAERHPEKYQGSSVSNLHVPCGNTTHASSLQHENSLLLTKDRMNV 300
Db 241 PSNNDLNTTEKRAAERHPEKYQGSSVSNLHVPCGNTTHASSLQHENSLLLTKDRMNV 300
Qy 301 KAFCNKSQKQGLARSQHNWAGSKETCNDRRTPSTTEKKYVDLNADPLCERKENWKQLPC 360
Db 301 KAFCNKSQKQGLARSQHNWAGSKETCNDRRTPSTTEKKYVDLNADPLCERKENWKQLPC 360
Qy 361 SENPRDTEVPWITLNSIOKVNEWFSRDELLGSDSDHGESNAKVADVDLNEVD 420
Db 361 SENPRDTEVPWITLNSIOKVNEWFSRDELLGSDSDHGESNAKVADVDLNEVD 420
Qy 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKASLPNLSHVTEN 480
Db 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKASLPNLSHVTEN 480
Qy 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVOKTPEMINQGTOTE 540
Db 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVOKTPEMINQGTOTE 540
Qy 541 QNGQVMNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTAEPITSSISNMELELNI 600
Db 541 QNGQVMNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTAEPITSSISNMELELNI 600
Qy 601 HNSKAPKKNLRKSTRHIALELVVSRNLSPNCTELOIDSCSSSEEEKKKYKNMPV 660
Db 601 HNSKAPKKNLRKSTRHIALELVVSRNLSPNCTELOIDSCSSSEEEKKKYKNMPV 660
Qy 661 RHSRNLQMEGKEPATGAKSKNKPNEQTSKRHSDTFPELKLTNAPGSFTKCSNTSELKE 720
Db 661 RHSRNLQMEGKEPATGAKSKNKPNEQTSKRHSDTFPELKLTNAPGSFTKCSNTSELKE 720
Qy 721 FVNPSPREEKEKLETVKYSNNAEDPKDMLSGERVLOQTERSVESSISLVPDGTGYTQ 780
Db 721 FVNPSPREEKEKLETVKYSNNAEDPKDMLSGERVLOQTERSVESSISLVPDGTGYTQ 780
Qy 781 ESISLLEVTSLGKAKTEPNKCVSOCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHEVNH 840
Db 781 ESISLLEVTSLGKAKTEPNKCVSOCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHEVNH 840
Qy 841 RETSIEMESELDAQYLQNTFKVSKRQSFALFSPNGNAEEECATFFSAHSGSLKKQSPKVT 900
Db 841 RETSIEMESELDAQYLQNTFKVSKRQSFALFSPNGNAEEECATFFSAHSGSLKKQSPKVT 900
Qy 901 FECEQKEENOGKNESTKPVQTVNITAGFPVVGOKPVDNAKCSIKGGRFCLSSQFRG 960
Db 901 FECEQKEENOGKNESTKPVQTVNITAGFPVVGOKPVDNAKCSIKGGRFCLSSQFRG 960

Qy 961 NETGLITPNKHGLQNPYRIPPIKSFVYKTKCKKNLLEENFEHMSPERMGNGENIP 1020
Db 961 NETGLITPNKHGLQNPYRIPPIKSFVYKTKCKKNLLEENFEHMSPERMGNGENIP 1020
Qy 1021 STVSTISRNNIRENVFKEASSNINEVGSSTNEVGSINEIGSSDENIOAELGRNRPKL 1080
Db 1021 STVSTISRNNIRENVFKEASSNINEVGSSTNEVGSINEIGSSDENIOAELGRNRPKL 1080
Qy 1081 NAMRLGVLPQPEYVKOSLPGSNCKHPEIKKOEYEEVVQVNTDFSPYLLSDNLEOPMGSS 1140
Db 1081 NAMRLGVLPQPEYVKOSLPGSNCKHPEIKKOEYEEVVQVNTDFSPYLLSDNLEOPMGSS 1140
Qy 1141 HASQVCSETPDDLLDGEIKEDTSFAENDIKESSAVFSKVQKGLSPSPFTTHLAQ 1200
Db 1141 HASQVCSETPDDLLDGEIKEDTSFAENDIKESSAVFSKVQKGLSPSPFTTHLAQ 1200
Qy 1201 GYRGAKKLESSEENISSEDEELPCFQHLLFGKVNIPSQSTRHSTVATECLSKNTEENL 1260
Db 1201 GYRGAKKLESSEENISSEDEELPCFQHLLFGKVNIPSQSTRHSTVATECLSKNTEENL 1260
Qy 1261 LSLKNSLDCSNVILAKASOEHLSEETKCSASLFSFSSOCSELEDLTANTNTQDPFLIGS 1320
Db 1261 LSLKNSLDCSNVILAKASOEHLSEETKCSASLFSFSSOCSELEDLTANTNTQDPFLIGS 1320
Qy 1321 SKQMRHQSOGVGLSDKELVSDDEERGTLGLENNOEQSMDSNLGEASGCESETSYSE 1380
Db 1321 SKQMRHQSOGVGLSDKELVSDDEERGTLGLENNOEQSMDSNLGEASGCESETSYSE 1380
Qy 1381 DCSGLSSQSDILITQORDTMOHNLKLOEMAELEAVLEQHSQSPSNSYPSIISSSALE 1440
Db 1381 DCSGLSSQSDILITQORDTMOHNLKLOEMAELEAVLEQHSQSPSNSYPSIISSSALE 1440
Qy 1441 DLNRPEOSTSEKAVLTQSSEYPISONPEGLSADKFEVSADSSTSKNKEPVERSSPSK 1500
Db 1441 DLNRPEOSTSEKAVLTQSSEYPISONPEGLSADKFEVSADSSTSKNKEPVERSSPSK 1500
Qy 1501 CPSLDRWYMHSCGSLQNRNYPQBELIKVVDVEEQLEESGPHDLTETSYLPRQDLEG 1560
Db 1501 CPSLDRWYMHSCGSLQNRNYPQBELIKVVDVEEQLEESGPHDLTETSYLPRQDLEG 1560
Qy 1561 TPYLESGISLFDPSDESDRAPSARVGNIPSTSTALKVPOLKVAESAQSPAAHHT 1620
Db 1561 TPYLESGISLFDPSDESDRAPSARVGNIPSTSTALKVPOLKVAESAQSPAAHHT 1620
Qy 1621 DTAGYNAMESVSREKPELTASTERNVKNRMSVMVSGLTPEEFMLVYKFKARKHHTLTNLI 1680
Db 1621 DTAGYNAMESVSREKPELTASTERNVKNRMSVMVSGLTPEEFMLVYKFKARKHHTLTNLI 1680
Qy 1681 TEETHVVMKTDAEFVCERTLYFLGIAGGKVVVSYFWVTQSIKERMKMLNEHDFEVRGDV 1740
Db 1681 TEETHVVMKTDAEFVCERTLYFLGIAGGKVVVSYFWVTQSIKERMKMLNEHDFEVRGDV 1740
Qy 1741 VNGRNHQPGRKRESODRIKFRGLETCGYPPTNPTDQLEHVMVQCGASVVKELSSFTL 1800
Db 1741 VNGRNHQPGRKRESODRIKFRGLETCGYPPTNPTDQLEHVMVQCGASVVKELSSFTL 1800
Qy 1801 GTGVHPVVVQPDWNTDNGFHAIGOMCEAPVVTREWVLDLSVALYQCOBELDYILIPQIPH 1860
Db 1801 GTGVHPVVVQPDWNTDNGFHAIGOMCEAPVVTREWVLDLSVALYQCOBELDYILIPQIPH 1860
Qy 1861 SHY 1863
Db 1861 SHY 1863
RESULT 3
ID AAB24219 standard; Protein; 1863 AA.
XX
AC AAB24219;
XX
XX 07-FEB-2001 (first entry)
DT
XX

DE Human BRCA1 (om13) protein sequence SEQ ID NO:6.
XX Human: BRCA1; chromosome 17; 17q21; breast cancer; ovarian cancer;
KW gene therapy; diagnosis; cytostatic; genetic susceptibility; mutation;
KW polymorphism; identification.
XX
OS Homo sapiens.
XX
PN US6130322-A.
PD 10-OCT-2000.
XX
XX 06-MAY-1998; 98US-0074476.
PF
PR 12-FEB-1996; 96US-0598591.
PR 12-DEC-1997; 97US-0798691.
XX
XX (GENE-) GENE LOGIC INC.
PI Zeng B, Thurber D, Olson SJ, Alvares CP, Allen ACP, Murphy PD;
PI Critz BS;
XX
DR WPI; 2000-646756/62.
DR N-PSDB; AAC60795.
XX
XX New coding sequence of the human BRCA1 gene, i.e. BRCA1 (om12), useful
PT in gene therapy, especially for preventing or treating breast or
PT ovarian cancer, as well as for diagnosing or monitoring breast or
PT ovarian cancer -
XX
PS
PS Example 4; Column 65-74; 56pp; English.
XX
CC AAC60793 to AAC60795 encode the human BRCA1 (om1-3) proteins given in
CC AAB24217 to AAB24219 respectively. BRCA1 is found on chromosome 17
CC mapping to position 17q21. The BRCA1 (om12) coding sequence is
CC specifically claimed in the present invention. The BRCA1 (om12) coding
CC sequence is useful in gene therapy, especially for preventing or treating
CC breast or ovarian cancer. It is also useful for diagnosing or monitoring
CC breast or ovarian cancer. Furthermore, the BRCA1 (om12) coding sequence
CC is useful for: (a) identifying individuals having BRCA1 gene mutations
CC and having an increased genetic susceptibility to breast or ovarian
CC cancer, or identifying a mutation that increases the genetic
CC susceptibility to breast or ovarian cancer; (b) avoiding
CC misinterpretation of polymorphisms found in the BRCA1 gene; (c)
CC determining the presence of a previously unknown mutation in the BRCA1
CC gene; (d) probing a human sample of the BRCA1 gene by allele to determine
CC the presence of either polymorphic alleles or mutations; and (e)
CC performing diagnosis with a reagent derived from the BRCA1 (om1) cDNA
CC sequence. AAC60796 to AAC60861 represent PCR primers for the BRCA1 gene,
CC which are used in an example from the present invention.
XX
SQ Sequence 1863 AA;

Query Match 100.0%; Score 9649; DB 21; Length 1863;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MDLSALRVEEVQNVINAMQKILECPICIELEIKPEYSTKCDHIFCKFCMLKLLNQKGPQSQ 60
Db 1 MDLSALRVEEVQNVINAMQKILECPICIELEIKPEYSTKCDHIFCKFCMLKLLNQKGPQSQ 60
Qy 61 CPLCKNDITKRSLOESTRFSOLVBEELKIIICAFQDLDTGLEVANSYNFAKKENNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRFSOLVBEELKIIICAFQDLDTGLEVANSYNFAKKENNSPEHLKD 120
Qy 121 EVSIIQSMGYRNRAKRLQSEPNPSLOETSLSVOLSNLGTVRTLTQRIOQKTSVYI 180
Db 121 EVSIIQSMGYRNRAKRLQSEPNPSLOETSLSVOLSNLGTVRTLTQRIOQKTSVYI 180
Qy 181 ELGSDSSEDTYKATYCSVGQDELLOITPQGTDRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDTYKATYCSVGQDELLOITPQGTDRDEISLDSAKKAACEFSETDVTNTEHHQ 240

QY 241 PSNNDLNTTEKRAAERPEKIQGSSVSNLHVPCGNTNTHASSLQHENSLSLLTKDRMNVE 300
 Db 241 PSNNDLNTTEKRAAERPEKIQGSSVSNLHVPCGNTNTHASSLQHENSLSLLTKDRMNVE 300
 QY 301 KAEFCNKSOPGLARSOHNWAGSKETCNDRRTPSTTEKKVLDLADPLCERKEWNKOLPC 360
 Db 301 KAEFCNKSOPGLARSOHNWAGSKETCNDRRTPSTTEKKVLDLADPLCERKEWNKOLPC 360
 QY 361 SENPROTEDEPWTITLNSSIQKVNWFSSRDELGSDSDHSGESESNAKVADVLVDLNEVD 420
 Db 361 SENPROTEDEPWTITLNSSIQKVNWFSSRDELGSDSDHSGESESNAKVADVLVDLNEVD 420
 QY 421 EYSGSEKIDLLASDPHEALICKSERVHKSVSFNIEDKIFGKYRKKASLPNLSHVTEN 480
 Db 421 EYSGSEKIDLLASDPHEALICKSERVHKSVSFNIEDKIFGKYRKKASLPNLSHVTEN 480
 QY 481 LIIGAFVTEPQIIIOERPLTNKLRKRPTSGLHPEDFIKKADLAVOKTPEMINQGTNOTE 540
 Db 481 LIIGAFVTEPQIIIOERPLTNKLRKRPTSGLHPEDFIKKADLAVOKTPEMINQGTNOTE 540
 QY 541 QNGQVMNITNSGHENKTGDSIQONEKNPNPIESLEKESAFKTKAEPISSISNMELELNI 600
 Db 541 QNGQVMNITNSGHENKTGDSIQONEKNPNPIESLEKESAFKTKAEPISSISNMELELNI 600
 QY 601 HNSKAPKKNRLRRKSTRHITHALELVVSRNLSPNCTELQIDSCSSSEEEKKKYNOMPV 660
 Db 601 HNSKAPKKNRLRRKSTRHITHALELVVSRNLSPNCTELQIDSCSSSEEEKKKYNOMPV 660
 QY 661 RHSRNLQMEGKEPATGAKSKSNPNQOTSKRHSDSDTFPELKLTPNAPGSKTCSNTSELKE 720
 Db 661 RHSRNLQMEGKEPATGAKSKSNPNQOTSKRHSDSDTFPELKLTPNAPGSKTCSNTSELKE 720
 QY 721 FVNPSLPREEKEKLETKVYKSNNAEDPKDMLSGSERVLOTERSVESSSISLVPCTDYGTO 780
 Db 721 FVNPSLPREEKEKLETKVYKSNNAEDPKDMLSGSERVLOTERSVESSSISLVPCTDYGTO 780
 QY 781 ESTSLLEVSTLGRKAKTEPNKVSQCAAFENPKGLIHGCSKDNNDTEGPKYPLGHEVNH 840
 Db 781 ESTSLLEVSTLGRKAKTEPNKVSQCAAFENPKGLIHGCSKDNNDTEGPKYPLGHEVNH 840
 QY 841 RETSIEMESSELDQYLONTFKVSKQSFALFSPNCAEECATFSAHSGSLKKQSPKVT 900
 Db 841 RETSIEMESSELDQYLONTFKVSKQSFALFSPNCAEECATFSAHSGSLKKQSPKVT 900
 QY 901 FECEKEENOGKNESNIKPVQTVNITAGFPVVGQKDPVDNAKCSIKGGSRFCLSQFRG 960
 Db 901 FECEKEENOGKNESNIKPVQTVNITAGFPVVGQKDPVDNAKCSIKGGSRFCLSQFRG 960
 QY 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVKTKCKKNLLEENPEEHSMSPERMGNENIP 1020
 Db 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVKTKCKKNLLEENPEEHSMSPERMGNENIP 1020
 QY 1021 STVSTISRNNIRENVFKEASSNINEVGSTNEVGSSINEIGSSDENIQAEGLRNGRPKL 1080
 Db 1021 STVSTISRNNIRENVFKEASSNINEVGSTNEVGSSINEIGSSDENIQAEGLRNGRPKL 1080
 QY 1081 NAMRLGVLPQPEVYKQSLPGSNCKHPEIKKQYEEVVQTVNTDFSPYLISDNLEQPMGSS 1140
 Db 1081 NAMRLGVLPQPEVYKQSLPGSNCKHPEIKKQYEEVVQTVNTDFSPYLISDNLEQPMGSS 1140
 QY 1141 HASQVCSPTDLDLDDGETIKEDTSPAENDIKESSAVFSKVQKGLSRSPSPHTHFLAQ 1200
 Db 1141 HASQVCSPTDLDLDDGETIKEDTSPAENDIKESSAVFSKVQKGLSRSPSPHTHFLAQ 1200
 QY 1201 GYBRGAKKLESSEENLSSDEELPCFQHLFGKVNIPQSTHRHSTVATECLSKNTEENL 1260
 Db 1201 GYBRGAKKLESSEENLSSDEELPCFQHLFGKVNIPQSTHRHSTVATECLSKNTEENL 1260
 QY 1261 LSLKNSLNCDSNOVILAKASQEHHLSEETKCSASLFSOCSSELEDLTANTQDPFLIGS 1320
 Db 1261 LSLKNSLNCDSNOVILAKASQEHHLSEETKCSASLFSOCSSELEDLTANTQDPFLIGS 1320
 QY 1321 SKOMRHQSSQVGLSKDELVSDEERGTCLENNQEQSMDSNLGEAASGCESETSVSE 1380

Db 1321 SKOMRHQSSQVGLSKDELVSDEERGTCLENNQEQSMDSNLGEAASGCESETSVSE 1380
 QY 1381 DCSGLSSQSDILTTQORDTMOHNLIKLOQEMAELEAVLEHQHQSOPNSYPSIISDSSALE 1440
 Db 1381 DCSGLSSQSDILTTQORDTMOHNLIKLOQEMAELEAVLEHQHQSOPNSYPSIISDSSALE 1440
 QY 1441 DLNRPQSTSEKAVLTQSSESSEYPISONPEGLSADKFEVSADSSTSKNKEPGVERSSPSK 1500
 Db 1441 DLNRPQSTSEKAVLTQSSESSEYPISONPEGLSADKFEVSADSSTSKNKEPGVERSSPSK 1500
 QY 1501 CPSLDDRWTMHSCSGSLQNRNYPSEELIKVVDVVEQQLEESGPHDLTETSYLPRODLEG 1560
 Db 1501 CPSLDDRWTMHSCSGSLQNRNYPSEELIKVVDVVEQQLEESGPHDLTETSYLPRODLEG 1560
 QY 1561 TPYLESISILFSDDPSPSEDRAPE SARVGNIPSPSTSAKVPQLKVAESAQSPAAAHFT 1620
 Db 1561 TPYLESISILFSDDPSPSEDRAPE SARVGNIPSPSTSAKVPQLKVAESAQSPAAAHFT 1620
 QY 1621 DTAGYNAMESVSREKPELTASTERNVKNRMSVMVSGLTPEEFMLVYKFARKKHITLTNLI 1680
 Db 1621 DTAGYNAMESVSREKPELTASTERNVKNRMSVMVSGLTPEEFMLVYKFARKKHITLTNLI 1680
 QY 1681 TEETTHVMKTDAEFVCERTLYFLGIAGGKVVVSYFWVTQSTIKERKMLNEHDFEVRGDV 1740
 Db 1681 TEETTHVMKTDAEFVCERTLYFLGIAGGKVVVSYFWVTQSTIKERKMLNEHDFEVRGDV 1740
 QY 1741 VNGRNHOGKPRARESODRKIFRGLEICCCYGPFTNMPTDQLEWMVQLCGVASVVKELSSFTL 1800
 Db 1741 VNGRNHOGKPRARESODRKIFRGLEICCCYGPFTNMPTDQLEWMVQLCGVASVVKELSSFTL 1800
 QY 1801 GTGVHPIVVVQPDAMTENDNGHAIGOMCEAPVVTREWLDSVALYOCQELDTYLIPOIPH 1860
 Db 1801 GTGVHPIVVVQPDAMTENDNGHAIGOMCEAPVVTREWLDSVALYOCQELDTYLIPOIPH 1860
 QY 1861 SHY 1863
 Db 1861 SHY 1863
 RESULT 4
 AAR97128
 ID AAR97128 standard; Protein; 1863 AA.
 XX
 AC AAR97128;
 XX
 DT 19-NOV-1996 (first entry)
 XX
 DE BRCA1, breast and ovarian cancer susceptibility gene product.
 XX
 KW BRCA1; breast cancer; ovary cancer; predisposing gene; diagnosis;
 KW susceptibility gene; prognosis; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO9605308-A1.
 PD 22-FEB-1996.
 XX
 PF 11-AUG-1995; 95WO-US10220.
 XX
 PR 07-JUN-1995; 95US-0488011.
 PR 12-AUG-1994; 94US-0289221.
 PR 02-SEP-1994; 94US-0300266.
 PR 16-SEP-1994; 94US-0308104.
 PR 29-NOV-1994; 94US-0348824.
 PR 24-MAR-1995; 95US-0409305.
 PR 07-JUN-1995; 95US-0483554.
 PR 07-JUN-1995; 95US-0487002.
 XX
 PA (MYRI-) MYRIAD GENETICS INC.
 PA (UTAH) UNIV UTAH RES FOUND.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Futreal PA, Goldgar DE, Harshman KD, Kamb A, Miki Y;
PI Shattuck-Eidens DM, Skolnick MH, Swenson J, Tavtigian SV;
PI Wiseman RW;
XX WPI; 1996-139704/14.
DR N-PSDB; AAT32601.
XX
PT New method for diagnosing a predisposition to breast and ovarian
PT cancer - by detecting a germline alteration in the BRCA1 gene or
PT gene regulatory sequence; for gene therapy and to screen for drugs
XX
PS Claim 12; Page 108-117; 200pp; English.
XX
XX This is the amino acid sequence of the protein encoded by the breast and
CC ovarian cancer susceptibility gene, BRCA1. Four kindred families
CC provided genetic evidence for localisation of BRCA1 to a sufficiently
CC small region for the application of positional cloning strategies. A
CC detailed map of transcripts was developed for the region of 17q21 between
CC D17S1321 and D17S1324. A combination of sequences obt'd. from cDNA
CC clones, hybrid-selected sequences and PCR prods. allowed construction of
CC a composite full-length BRCA1 cDNA (see AAT32612 for genomic sequence).
CC The BRCA1 cDNA revealed an open reading frame encoding a protein of
CC 208 kD. The protein also contains a sequence near the amino terminus
CC which had homology to zinc finger domains, esp. the sequence contains
CC Cys and His residues present as a C3HC4 zinc finger consensus motif.
CC The isolated cDNA is used in methods for either diagnosis of the
CC predisposition to cancer (partic. breast and ovarian cancer), or for the
CC diagnosis or prognosis of cancer, and also in gene-based therapies
CC directed at cancer cells.
XX
SQ Sequence 1863 AA;

Query Match 99.9%; Score 9642; DB 17; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFVCLLNQKGPSP 60
DB 1 MDSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFVCLLNQKGPSP 60

QY 61 CPLCKNDITKRSLOESTRFSOLVBEELLKICAFOLDTGLEYANSYNFAKKENNSPEHLKD 120
DB 61 CPLCKNDITKRSLOESTRFSOLVBEELLKICAFOLDTGLEYANSYNFAKKENNSPEHLKD 120

QY 121 EYSIIQSMGYRNRRARLLQSEPNPSLOETSLSVOLSNLGTVRTLRKQIOPKTSVYI 180
DB 121 EYSIIQSMGYRNRRARLLQSEPNPSLOETSLSVOLSNLGTVRTLRKQIOPKTSVYI 180

QY 181 ELGSDSSEDTYNKATYCSVGDQELLIQITPQGTDRDEISLDSAKKAACEFSETDVTNTEHHQ 240
DB 181 ELGSDSSEDTYNKATYCSVGDQELLIQITPQGTDRDEISLDSAKKAACEFSETDVTNTEHHQ 240

QY 241 PSNNDLNTEKRAARHPPEKPYGGSVSNLHVPCGTNTHASSLQHNSSLLLTAKDRNVE 300
DB 241 PSNNDLNTEKRAARHPPEKPYGGSVSNLHVPCGTNTHASSLQHNSSLLLTAKDRNVE 300

QY 301 KAEFCNKSQKPLARSOHNRWAGSKETCNDRTPTSEKKVDLNADPLCERKENWKKLPC 360
DB 301 KAEFCNKSQKPLARSOHNRWAGSKETCNDRTPTSEKKVDLNADPLCERKENWKKLPC 360

QY 361 SENPRDTEDPWITLNSSIQKVNWFWSRDELGSDSDSHDGESESNKAVADVLDVLNEVD 420
DB 361 SENPRDTEDPWITLNSSIQKVNWFWSRDELGSDSDSHDGESESNKAVADVLDVLNEVD 420

QY 421 EYSGSSEKIDLLASDPHALCKSERVHKSVSNEIEDKIFGKYTYRKASLPNLVSHVTEN 480
DB 421 EYSGSSEKIDLLASDPHALCKSERVHKSVSNEIEDKIFGKYTYRKASLPNLVSHVTEN 480

QY 481 LIIGAFVTEPQIOPRLTNLKKRRTSLGLHPEDFLKKADLAVOKTPEMINOGTNOTE 540
DB 481 LIIGAFVTEPQIOPRLTNLKKRRTSLGLHPEDFLKKADLAVOKTPEMINOGTNOTE 540

QY 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPTESLEKESAFKTKAEPISISSINMELELNI 600
DB 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPTESLEKESAFKTKAEPISISSINMELELNI 600

QY 601 HNSKAPKKNRLRRKSSRTHIALELVVSRNLSPNCTELQIDSCSSSEEEIKKKYQNMPV 660
DB 601 HNSKAPKKNRLRRKSSRTHIALELVVSRNLSPNCTELQIDSCSSSEEEIKKKYQNMPV 660

QY 661 RHSNLIOLMEGKEPATGAKKSNKNEOTSKRHDSDTPELKLTNAPGFTKCSNTSELKE 720
DB 661 RHSNLIOLMEGKEPATGAKKSNKNEOTSKRHDSDTPELKLTNAPGFTKCSNTSELKE 720

QY 721 FVNPSPREEKEEKLTVKVSNNNAEDPKDMLSGSERVLQTERSVESSISLVPGDYGTQ 780
DB 721 FVNPSPREEKEEKLTVKVSNNNAEDPKDMLSGSERVLQTERSVESSISLVPGDYGTQ 780

QY 781 ESTSLLEVSTLGKAKTPNKCVCSCAAAFENPKGLIHGCSKDNRNDTEGFKYPLGHEVNH 840
DB 781 ESTSLLEVSTLGKAKTPNKCVCSCAAAFENPKGLIHGCSKDNRNDTEGFKYPLGHEVNH 840

QY 841 RETSIEMEESFELDAQYLQNTFKVSKROSFALFSPNGNAEECATFSAHSGSLKKQSPKVT 900
DB 841 RETSIEMEESFELDAQYLQNTFKVSKROSFALFSPNGNAEECATFSAHSGSLKKQSPKVT 900

QY 901 FECQKEENOGKNESNIKPVQTVNITAGFPVVGQKDPVDNAKCSIKGSRFCLSSQFRG 960
DB 901 FECQKEENOGKNESNIKPVQTVNITAGFPVVGQKDPVDNAKCSIKGSRFCLSSQFRG 960

QY 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVKTKCKNLEENFEHSHMSPEREMGENIP 1020
DB 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVKTKCKNLEENFEHSHMSPEREMGENIP 1020

QY 1021 STVSTISRNINRENVFKEASSNINEVGSSSTNEIGSSSDENIQAEIAGNRGPKL 1080
DB 1021 STVSTISRNINRENVFKEASSNINEVGSSSTNEIGSSSDENIQAEIAGNRGPKL 1080

QY 1081 NAMRLGLVLAPEVYKQSLPGSNCKHPETKQYEEVQTVNTDFSPYILSDNLEQPMGSS 1140
DB 1081 NAMRLGLVLAPEVYKQSLPGSNCKHPETKQYEEVQTVNTDFSPYILSDNLEQPMGSS 1140

QY 1141 HASQVCESETPDDLDDGEIKEDETSFAENDIKESSAVFSKSVQKGLSRSPSPFTTHLAQ 1200
DB 1141 HASQVCESETPDDLDDGEIKEDETSFAENDIKESSAVFSKSVQKGLSRSPSPFTTHLAQ 1200

QY 1201 GYRGAKKLESSEENLSSEDEELPCFOHLLFGKVNYPISQSTRHSTVATECLSKNTEENL 1260
DB 1201 GYRGAKKLESSEENLSSEDEELPCFOHLLFGKVNYPISQSTRHSTVATECLSKNTEENL 1260

QY 1261 LSLKNSLNDCSNQVILAKASQEHLSSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS 1320
DB 1261 LSLKNSLNDCSNQVILAKASQEHLSSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS 1320

QY 1321 SKOMRHOSQSGVGLSDKELVSDDEERGTLGLENNQBEQSMDSNLGEAASCESTSVSE 1380
DB 1321 SKOMRHOSQSGVGLSDKELVSDDEERGTLGLENNQBEQSMDSNLGEAASCESTSVSE 1380

QY 1381 DCSGLSSQSDILTQQORDTQOHNLIKLQOEAEVLEQHGSOPSNSYPSTISDSSALE 1440
DB 1381 DCSGLSSQSDILTQQORDTQOHNLIKLQOEAEVLEQHGSOPSNSYPSTISDSSALE 1440

QY 1441 DLNRPEQSTSEKAVLTQSSEYPISONPEGLSADKFEVSADSDSTSKNKEPVERSSPSK 1500
DB 1441 DLNRPEQSTSEKAVLTQSSEYPISONPEGLSADKFEVSADSDSTSKNKEPVERSSPSK 1500

QY 1501 CPSLDDRWYHSCGSLQNRNYPQBELIKVVDVEBOOLESGPHDLTETSYLPRQDLEG 1560
DB 1501 CPSLDDRWYHSCGSLQNRNYPQBELIKVVDVEBOOLESGPHDLTETSYLPRQDLEG 1560

QY 1561 TPYLESGISLFSDDPSPEDRAPESARVGNIPSSTSALKVPOLKVAESAQSPAAHTT 1620
DB 1561 TPYLESGISLFSDDPSPEDRAPESARVGNIPSSTSALKVPOLKVAESAQSPAAHTT 1620

QY 1621 DTAGYNAMEESVREKPELTAETERVNRKRMVMVSGLTPEEFMLVYKFAKHHITLTNLI 1680

||||| 1621 DTAGYNAMESVSREKPELTASTERVNKRMSMVVSGLTPEEFMLVYKFARKHHITLNL 1680
QY 1681 TETTHVVMKTDAEFCERTLKAYELGIAGKVVVSVFWVTQSTKERKMLNEHDFEVRGDV 1740
Db 1681 TETTHVVMKTDAEFCERTLKAYELGIAGKVVVSVFWVTQSTKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHQGPKRARESQRDRIFRGLGICCYGPTNMPDQLEWVQLCGASVVKELSSFTL 1800
Db 1741 VNGRNHQGPKRARESQRDRIFRGLGICCYGPTNMPDQLEWVQLCGASVVKELSSFTL 1800
QY 1801 GTGVHPVIVVQPDADWETDNGFHAIGMCBAPVVTREWVLDVALYQCQELDYLIPIQIPH 1860
Db 1801 GTGVHPVIVVQPDADWETDNGFHAIGMCBAPVVTREWVLDVALYQCQELDYLIPIQIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863
|||
RESULT 5
ID AAR81481 standard; Protein; 1863 AA.
XX AAR81481;
AC AAR81481;
DT 30-SEP-1996 (first entry)
XX BRCA1.
XX
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
XX
XX Homo sapiens.
XX WO9605306-A2.
PN
XX
XX 22-FEB-1996.
XX
XX 11-AUG-1995; 95WO-US10202.
XX
PR 07-JUN-1995; 95US-0483553.
PR 12-AUG-1994; 94US-0289221.
PR 02-SEP-1994; 94US-0300266.
PR 16-SEP-1994; 94US-0308104.
PR 29-NOV-1994; 94US-0348824.
PR 24-MAR-1995; 95US-0409305.
PR 07-JUN-1995; 95US-0480784.
XX
XX (MYRI-) MYRIAD GENETICS INC.
XX (CANC-) CANCER INST.
XX (RECH-) CENT RECH DU CHUL.
XX
XX Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
PI WPI; 1996-139702/14.
XX N-PSDB; AAT17438.
DR
XX
XX New nucleic acid and polypeptide for mutant or polymorphic BRCA1
PT gene - for diagnosis and therapy of human breast and ovarian cancer
PT and for diagnosing pre-disposition to these cancers
XX
PS Claim 1; Page 119-128; 218pp; English.
XX
CC This sequence represents the protein encoded by the human breast and
CC ovarian cancer predisposing gene (BRCA1). Mutations of this sequence
CC (see AAR81483-R81497 and AAR81499-R81546) can be used as immunogens for
CC antibody production. The mutant BRCA1 genes have at least 1 mutation or
CC polymorphism in comparison to the cDNA encoding this sequence. By
CC detecting a germline alteration in this gene, a predisposition for
CC breast and ovarian cancer can be diagnosed. In one method, BRCA1 mRNA
CC isolated from a tissue sample from a subject has a probe, corresponding

.CC to a fragment of the cDNA encoding this sequence (or an allele-specific
CC probe for a mutation of it sequence), added to it. The conditions allow
CC for hybridisation of the probe to the mRNA, and any hybridisation which
CC occurs is detected. Alternatively the BRCA1 gene in the tissue sample
CC is isolated, and a shift in electrophoretic mobility of single stranded
CC DNA from the sample on a non-denaturing polyacrylamide gel indicates a
CC mutation. These methods of detection can also diagnose a lesion
CC neoplasia associated with the BRCA1 locus. The methods may be used in
CC gene therapy, protein replacement therapy and protein mimetics, and may
CC be used to screen for drugs in cancer therapy.
XX
SQ Sequence 1863 AA;
Query Match 99.9%; Score 9642; DB 17; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MDLSALRVEEVQNVINAMOKILECPICLEIKPEVSTKCDHIFCKFCMLKLLNQKGPQSQ 60
Db 1 MDLSALRVEEVQNVINAMOKILECPICLEIKPEVSTKCDHIFCKFCMLKLLNQKGPQSQ 60
QY 61 CPLCKNDITKRSLOESTRFSQVLELLKIIICAFQDLDTGLEYANSYNFAKKENNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRFSQVLELLKIIICAFQDLDTGLEYANSYNFAKKENNSPEHLKD 120
QY 121 EYSIIQSMGYRNRAKRLQSEPNPSLOETSLSVQLSNLGTVRTLTQRKRIOPQKTSVYI 180
Db 121 EYSIIQSMGYRNRAKRLQSEPNPSLOETSLSVQLSNLGTVRTLTQRKRIOPQKTSVYI 180
QY 181 ELGSDSSEDTVNKATYCVGDQELQITPQGRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDTVNKATYCVGDQELQITPQGRDEISLDSAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNNDLNTTEKRAAERHPEKYQSSVSNLHVPCGTNTHASSLOHENSLLLTDKRMNVE 300
Db 241 PSNNDLNTTEKRAAERHPEKYQSSVSNLHVPCGTNTHASSLOHENSLLLTDKRMNVE 300
QY 301 KAEFCNKSQPGLARSOHNRWAGSKETCNDRRTPSTPEKKVVDLNADPLCERKENWKQKLPFC 360
Db 301 KAEFCNKSQPGLARSOHNRWAGSKETCNDRRTPSTPEKKVVDLNADPLCERKENWKQKLPFC 360
QY 361 SENPRDTEVPWITLNSSTQKYNWFPSRDELLGSDSDHDSSESNAKVADVLVDLNEVD 420
Db 361 SENPRDTEVPWITLNSSTQKYNWFPSRDELLGSDSDHDSSESNAKVADVLVDLNEVD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKKASLNLNLSHTEN 480
Db 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKKASLNLNLSHTEN 480
QY 481 LIIGAPVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVOKTPEMINOQTNQTE 540
Db 481 LIIGAPVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVOKTPEMINOQTNQTE 540
QY 541 QNGQVMNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISISSISNMELELNI 600
Db 541 QNGQVMNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISISSISNMELELNI 600
QY 601 HNSKAPKKNLRLRKSSTRHIALELVVSRNLSPPPNCTELOIDSCSSSEIKKKKYNQMPV 660
Db 601 HNSKAPKKNLRLRKSSTRHIALELVVSRNLSPPPNCTELOIDSCSSSEIKKKKYNQMPV 660
QY 661 RUSRNQLMEGKEPATGAKKSNKPNQTSKRHDSDTFFPELKTNPAGSFTKCSNTSELKE 720
Db 661 RUSRNQLMEGKEPATGAKKSNKPNQTSKRHDSDTFFPELKTNPAGSFTKCSNTSELKE 720
QY 721 FVNPSLPREEKEKLETVKVSNNAEDPKDMLSGERVLQTERSVESSISLVPDGYGTQ 780
Db 721 FVNPSLPREEKEKLETVKVSNNAEDPKDMLSGERVLQTERSVESSISLVPDGYGTQ 780
QY 781 ESISLLEVTLGAKTEPNKCVSQCAAFENPKGLIHGCSKDNRNDETEGPKYPLGHEVNH 840
Db 781 ESISLLEVTLGAKTEPNKCVSQCAAFENPKGLIHGCSKDNRNDETEGPKYPLGHEVNH 840

QY	181	ELGSDSSEDTVVKATYCSVGQDQLLIQTQGTREDEISLSAKAAACEFSETDVTNTEHHQ	240
Db	181	ELGSDSSEDTVVKATYCSVGQDQLLIQTQGTREDEISLSAKAAACEFSETDVTNTEHHQ	240
QY	241	PSNNDLNTTEKRAAERHPKQYQSSVSNLHVPCGCTNTIASSLQHNESLLLTQKDRMNVE	300
Db	241	PSNNDLNTTEKRAAERHPKQYQSSVSNLHVPCGCTNTIASSLQHNESLLLTQKDRMNVE	300
QY	301	KAEFCKNSKQPLGARSQHRNWRAGSKETCNDRRPTSPTEKKVVDLNADPLCERKEWNKOKLPC	360
Db	301	KAEFCKNSKQPLGARSQHRNWRAGSKETCNDRRPTSPTEKKVVDLNADPLCERKEWNKOKLPC	360
QY	361	SENPRDTEVPWITLNSSIQKYNWFSRSDDELIGSDSDHDSGESENAKVADVLVDLNEVD	420
Db	361	SENPRDTEVPWITLNSSIQKYNWFSRSDDELIGSDSDHDSGESENAKVADVLVDLNEVD	420
QY	421	EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDIKFKTYRKKASLPLNSHVTE	480
Db	421	EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDIKFKTYRKKASLPLNSHVTE	480
QY	481	LIIGAFTVEPQIIQERPLNKLKRKRPTSGLHPEDFIKKADLAVOKTPEMINOGTNOTE	540
Db	481	LIIGAFTVEPQIIQERPLNKLKRKRPTSGLHPEDFIKKADLAVOKTPEMINOGTNOTE	540
QY	541	ONGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINMELELNI	600
Db	541	ONGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINMELELNI	600
QY	601	HNSKAPKKNLRRKSSTRIHALELVVSRNLSPPNCTELQIDSCSSSEIEIKKKYNQMPV	660
Db	601	HNSKAPKKNLRRKSSTRIHALELVVSRNLSPPNCTELQIDSCSSSEIEIKKKYNQMPV	660
QY	661	RHSRNQLMEGKEPATGAKSKSNPEQTSKRHDSDFPELKLITNAPGCTFKCSNTSELKE	720
Db	661	RHSRNQLMEGKEPATGAKSKSNPEQTSKRHDSDFPELKLITNAPGCTFKCSNTSELKE	720
QY	721	FVNPSLPREEKEEKLTVKVSNNAEADPKDMLLSGERVLQTERSVESSISLVPGTDYGTQ	780
Db	721	FVNPSLPREEKEEKLTVKVSNNAEADPKDMLLSGERVLQTERSVESSISLVPGTDYGTQ	780
QY	781	ESISLLEVSTLGAKEPKNKYSQCAAFENPKGLIHGCSKDRNDTEGFKYPLGHEVNHIS	840
Db	781	ESISLLEVSTLGAKEPKNKYSQCAAFENPKGLIHGCSKDRNDTEGFKYPLGHEVNHIS	840
QY	841	RETSTEMESELDAQYLQNTFKVSKROSALFSPNPGNAEECATFSAHSGSLKKQSPKVT	900
Db	841	RETSTEMESELDAQYLQNTFKVSKROSALFSPNPGNAEECATFSAHSGSLKKQSPKVT	900
QY	901	FCEQKEENOGKNESNIKPVQTVNITAGPPVVGQDKPVDNAKCSIKGGSRFCLSSQFRG	960
Db	901	FCEQKEENOGKNESNIKPVQTVNITAGPPVVGQDKPVDNAKCSIKGGSRFCLSSQFRG	960
QY	961	NETGLITPNKHGLLQNPYRIPPLFKTSFVKTKCKNLLNEENFEHSMSPEREMGNENIP	1020
Db	961	NETGLITPNKHGLLQNPYRIPPLFKTSFVKTKCKNLLNEENFEHSMSPEREMGNENIP	1020
QY	1021	STVSTISRNNIRENVFKEASSNINEVGSTNEVGSSINEIGSSDENIOAELGRNRGPKL	1080
Db	1021	STVSTISRNNIRENVFKEASSNINEVGSTNEVGSSINEIGSSDENIOAELGRNRGPKL	1080
QY	1081	NAMLRLGVLOPEVYKQSLPGSNCKHPEIKKQEEYEVVQTVNTDFSPYLLSDNLEQPMGSS	1140
Db	1081	NAMLRLGVLOPEVYKQSLPGSNCKHPEIKKQEEYEVVQTVNTDFSPYLLSDNLEQPMGSS	1140
QY	1141	HASQVCSETPDDLDDGETIKEDTSAENDIKESSAVFSKVQKGLSRSPSFTHTHLAQ	1200
Db	1141	HASQVCSETPDDLDDGETIKEDTSAENDIKESSAVFSKVQKGLSRSPSFTHTHLAQ	1200
QY	1201	GYRRGAKKLLESSEENLSSDEELPCFOHLLFGKVNNIPSQSTRHSTVATECLSKNTEENL	1260
Db	1201	GYRRGAKKLLESSEENLSSDEELPCFOHLLFGKVNNIPSQSTRHSTVATECLSKNTEENL	1260

QY	1261	LSLKNSLNCDSNQVILAKAQEHLSSEETKCSASLFSQCSSELEDLTANTNQDPELIGS	1320
Db	1261	LSLKNSLNCDSNQVILAKAQEHLSSEETKCSASLFSQCSSELEDLTANTNQDPELIGS	1320
QY	1321	SKQMRHQSESQGVGLSDKELVSDDEERGTLGLENNOEBOGSMDSNLGEAAGCESETSVSE	1380
Db	1321	SKQMRHQSESQGVGLSDKELVSDDEERGTLGLENNOEBOGSMDSNLGEAAGCESETSVSE	1380
QY	1381	DCSGLSSQSDILTTQQRDTHQNLIKLOQEMAELEAVLQHQGSQPSNSVPSIISDSASLE	1440
Db	1381	DCSGLSSQSDILTTQQRDTHQNLIKLOQEMAELEAVLQHQGSQPSNSVPSIISDSASLE	1440
QY	1441	DLRNPQSTSEKAVLTSQKSSEYPIISQNPESGLSADKFEVSADSTSKNKEPGVERSSPSK	1500
Db	1441	DLRNPQSTSEKAVLTSQKSSEYPIISQNPESGLSADKFEVSADSTSKNKEPGVERSSPSK	1500
QY	1501	CPSLDRRWYHSCSGSLQNRNYPQSEELIKVVDVVEQQLEESGPHDLTETSYLPRODLEG	1560
Db	1501	CPSLDRRWYHSCSGSLQNRNYPQSEELIKVVDVVEQQLEESGPHDLTETSYLPRODLEG	1560
QY	1561	TPYLESIGLSLSDPESDSEDRAPEARVGNIPSPSTSALKVPOLKVAESAQSPAAAHIT	1620
Db	1561	TPYLESIGLSLSDPESDSEDRAPEARVGNIPSPSTSALKVPOLKVAESAQSPAAAHIT	1620
QY	1621	DRAGYNAMESVSREKPELTASTERVNKRMSVYVSGLTPEEFMLVYKFARKHHITLTNLI	1680
Db	1621	DRAGYNAMESVSREKPELTASTERVNKRMSVYVSGLTPEEFMLVYKFARKHHITLTNLI	1680
QY	1681	TEETTHVVKMTDAEFVCERTLKYFLGIAGGKVVVYFWVTQSIKERKMLNEHDFEVRGDV	1740
Db	1681	TEETTHVVKMTDAEFVCERTLKYFLGIAGGKVVVYFWVTQSIKERKMLNEHDFEVRGDV	1740
QY	1741	VNGRNHOGPKRARESODRKIFRGLEICCYGPTNMPDTQLEWVQLCGASVVKELSSFTL	1800
Db	1741	VNGRNHOGPKRARESODRKIFRGLEICCYGPTNMPDTQLEWVQLCGASVVKELSSFTL	1800
QY	1801	GTGVHPVIVVQPDAMPTDNGFHAIGQCEAPVVTREWLDSVALYQCQELDTYLIQPIPH	1860
Db	1801	GTGVHPVIVVQPDAMPTDNGFHAIGQCEAPVVTREWLDSVALYQCQELDTYLIQPIPH	1860
QY	1861	SHY 1863	
Db	1861	SHY 1863	
RESULT 7			
AAY32033			
ID	AAY32033 standard; protein; 1863 AA.		
XX			
AC	AAY32033;		
XX			
DT	05-JAN-2000 (first entry)		
XX			
DE	Human BRCA1 protein.		
XX			
KW	BRCA1; p53 protein; p21 gene; human; tumour suppressor;		
KW	transcriptional activator; breast cancer; cell proliferation;		
KW	apoptosis; diagnosis; anticancer; antitumour; drug screening.		
XX			
OS	Homo sapiens.		
XX			
FH	Key		
FT	Binding-site		
FT	/note= "binding site of p53 protein"		
FT	Region		
FT	/note= "NLS region"		
XX			
PN	WO9950280-A1.		
XX			
PD	07-OCT-1999.		
XX			
PF	31-MAR-1999; 99WO-US07150.		
XX			

PR 31-MAR-1998; 98US-0080146.
XX (UYPE-) UNIV PENNSYLVANIA.
XX El-Deiry WS, Weber BL;
XX WPI; 1999-601319/51.
XX Nucleic acid involved in BRCA-1-mediated control of transcriptional
PT regulation of tumour suppressor genes and related peptides, used to
PT screen for modulators for use as anticancer agents -
XX
PS Claim 9; Page 76-80; 94pp; English.
XX This sequence represents the human BRCA1 protein. BRCA-1
CC inhibits S-phase cell-cycle progression by transactivating
CC expression of p21 in a p53-independent manner. It also
CC interacts with p53 (see AA32034) in vitro and in vivo, and
CC increases p53-dependent transcription from the p21 (see AA220131)
CC and bax promoters. These newly discovered pathways of BRCA1 action
CC provide novel targets to which pharmaceutical agents capable of
CC affecting cell proliferation and apoptosis can be isolated, and
CC can then be used in the treatment and control of cellular
CC proliferation disorders, such as breast cancer. Assays and
CC compositions for identifying compounds that enhance or repress
CC cellular proliferation via these BRCA1-mediated pathways are
CC disclosed.
XX
SQ Sequence 1863 AA;
Query Match 99.9%; Score 9642; DB 20; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MDLSALRVEEVQNVINAMQKILECPICILELKEPVSFKCDHIFKFCMLKLLNKGKPSQ 60
DB 1 MDLSALRVEEVQNVINAMQKILECPICILELKEPVSFKCDHIFKFCMLKLLNKGKPSQ 60
QY 61 CPLCNDITKRSLOESTRFSOLVPELLKIIICAFOLDGTGLEVANSYNFAKKENNSPEHLKD 120
DB 61 CPLCNDITKRSLOESTRFSOLVPELLKIIICAFOLDGTGLEVANSYNFAKKENNSPEHLKD 120
QY 121 EYSIIQSGMYRNRARLLQSEPNPSLQETSLSVQLSNLGTVRTLRTKQRIQPKTSVYI 180
DB 121 EYSIIQSGMYRNRARLLQSEPNPSLQETSLSVQLSNLGTVRTLRTKQRIQPKTSVYI 180
QY 181 ELGSDSSEDTVNKATYCSVGQDQLLOITPQGTREISLDSAKKAACEFSETDVTNTEHHQ 240
DB 181 ELGSDSSEDTVNKATYCSVGQDQLLOITPQGTREISLDSAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNNDLNTTEKRAAERHPEKYQGSVSNLHVPCGTNTHASSLQHNSSLLLTKDRMNV 300
DB 241 PSNNDLNTTEKRAAERHPEKYQGSVSNLHVPCGTNTHASSLQHNSSLLLTKDRMNV 300
QY 301 KAEFCNKSQPGELARSQHRNAGSKETCNDRRTPSTEKKVDLADPLCEKERNKQKLPC 360
DB 301 KAEFCNKSQPGELARSQHRNAGSKETCNDRRTPSTEKKVDLADPLCEKERNKQKLPC 360
QY 361 SENPRDTEVPWITLNSIQKVNWFSSRDELGSDSDHGESESNAKVADVLVNEVD 420
DB 361 SENPRDTEVPWITLNSIQKVNWFSSRDELGSDSDHGESESNAKVADVLVNEVD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKYRKKASLPNLSHVTEN 480
DB 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKYRKKASLPNLSHVTEN 480
QY 481 LIIGAFVTEPQIIQERPLTNLKRKRRTSGLHPDEFIKKADLAVQKTPEMINGTNOTE 540
DB 481 LIIGAFVTEPQIIQERPLTNLKRKRRTSGLHPDEFIKKADLAVQKTPEMINGTNOTE 540
QY 541 QNGQVNNITNSCHENKTKGDSIQENKPNPIESLEKESAFKTKAEPISSSISNMELEINI 600
DB 541 QNGQVNNITNSCHENKTKGDSIQENKPNPIESLEKESAFKTKAEPISSSISNMELEINI 600

QY 601 HNSKAPKKNRLRRKSSTRHIIHALELVVSRNLSPNCTELOIDSCSSSEIKKKKKYNOMPV 660
DB 601 HNSKAPKKNRLRRKSSTRHIIHALELVVSRNLSPNCTELOIDSCSSSEIKKKKKYNOMPV 660
QY 661 RHSRNLQLMGKKEPATGAKKSNKPNEQTSKRHSDTFFPELKLITNAPGSFTKCSNTSELKE 720
DB 661 RHSRNLQLMGKKEPATGAKKSNKPNEQTSKRHSDTFFPELKLITNAPGSFTKCSNTSELKE 720
QY 721 FVNPSLPREKEEKLETVKVSNNAEDPKDMLSGSERVLQTSERSVSSSISLVPGYDTGTQ 780
DB 721 FVNPSLPREKEEKLETVKVSNNAEDPKDMLSGSERVLQTSERSVSSSISLVPGYDTGTQ 780
QY 781 ESIISLLEVSTLGGAKTPEPKVCVSOCAAFENPKGLTHGCSKDNNDTEGFKYPLGHEVNH 840
DB 781 ESIISLLEVSTLGGAKTPEPKVCVSOCAAFENPKGLTHGCSKDNNDTEGFKYPLGHEVNH 840
QY 841 RETSIEMEESLDAQYLQNTFKVSKROSFAFSPNGNAEEECATFSAHSGSLKKQSPKV 900
DB 841 RETSIEMEESLDAQYLQNTFKVSKROSFAFSPNGNAEEECATFSAHSGSLKKQSPKV 900
QY 901 FECQKEENQKKNESNKPVTVNITAGFPVVGOKDKPVDNAKCSIKGGSRFCLSQPRG 960
DB 901 FECQKEENQKKNESNKPVTVNITAGFPVVGOKDKPVDNAKCSIKGGSRFCLSQPRG 960
QY 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVTKCKKNLLEENFEHSHSPEREMGENIP 1020
DB 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVTKCKKNLLEENFEHSHSPEREMGENIP 1020
QY 1021 STVSTISRNIRENVFKEASSNINEVGSSNINEVGSSINEIGSSDENIQAEILGRNRGPKL 1080
DB 1021 STVSTISRNIRENVFKEASSNINEVGSSNINEVGSSINEIGSSDENIQAEILGRNRGPKL 1080
QY 1081 NAMRLGLVLOPEVYKQSLPGSNCKHPEIKOEYEVQTVNTDFSPYLISNLEOPMGSS 1140
DB 1081 NAMRLGLVLOPEVYKQSLPGSNCKHPEIKOEYEVQTVNTDFSPYLISNLEOPMGSS 1140
QY 1141 HASQVCSSETPDDLDDGEIKEEDTSAENDIKESSAVFSKVQKGLSRSPSPFTHTHQAQ 1200
DB 1141 HASQVCSSETPDDLDDGEIKEEDTSAENDIKESSAVFSKVQKGLSRSPSPFTHTHQAQ 1200
QY 1201 GYRGAKKLESSEENLSEDEELPCFQHLFLGKVNINPSQSTRHSTVATECLSKNTEENL 1260
DB 1201 GYRGAKKLESSEENLSEDEELPCFQHLFLGKVNINPSQSTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDSCSNQVILAKASQEHHLSEETKCSASFSSQCSSELEDLTANTNTQDPFLTGS 1320
DB 1261 LSLKNSLNDSCSNQVILAKASQEHHLSEETKCSASFSSQCSSELEDLTANTNTQDPFLTGS 1320
QY 1321 SKOMRHQSESQGVGLSDKELVSDDEERGTLGEENQEBOSMDSNLGEAASCESTSVSE 1380
DB 1321 SKOMRHQSESQGVGLSDKELVSDDEERGTLGEENQEBOSMDSNLGEAASCESTSVSE 1380
QY 1381 DCSGLSSQSDILTTQQRDTMOHNLIKLQOEMAELEAVLEQHGSQPSNSYPSIISDSSALE 1440
DB 1381 DCSGLSSQSDILTTQQRDTMOHNLIKLQOEMAELEAVLEQHGSQPSNSYPSIISDSSALE 1440
QY 1441 DLNRPEOSTSEKAVLTQSKSSEYPISONPEGLSADKFEVSADSTSKNKEPGEVERSSPSK 1500
DB 1441 DLNRPEOSTSEKAVLTQSKSSEYPISONPEGLSADKFEVSADSTSKNKEPGEVERSSPSK 1500
QY 1501 CPSLDDRWYHSCSGSLQNRNYPQOEELIKVDVEEQOLEESGPHDLTETSYLPQDLEG 1560
DB 1501 CPSLDDRWYHSCSGSLQNRNYPQOEELIKVDVEEQOLEESGPHDLTETSYLPQDLEG 1560
QY 1561 TPYLESGISLFSDDPESDPEDRAPESARVGNIPSSTSALKVPOLKVAESAQSPAHAHTT 1620
DB 1561 TPYLESGISLFSDDPESDPEDRAPESARVGNIPSSTSALKVPOLKVAESAQSPAHAHTT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVYSGLTPEEFMLVYKFAKHHTLTNLI 1680
DB 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVYSGLTPEEFMLVYKFAKHHTLTNLI 1680

QY 1681 TEETHVVMKTDABFVCERTILKYFLGIAGGKVVVSYFWVTQSTIKERKMLNEHDFEVRGDV 1740
Db 1681 TEETHVVMKTDABFVCERTILKYFLGIAGGKVVVSYFWVTQSTIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHOGPKRARESQRKIFRGLIEICYGPFTNMPTDQLEWVQLCGASVVKELSSFTL 1800
Db 1741 VNGRNHOGPKRARESQRKIFRGLIEICYGPFTNMPTDQLEWVQLCGASVVKELSSFTL 1800
QY 1801 GTGVHPITVVVQPDWEDNGFHAIGOMCEAPVVTREWVLDSVALYQCQELDTYLIPQIPH 1860
Db 1801 GTGVHPITVVVQPDWEDNGFHAIGOMCEAPVVTREWVLDSVALYQCQELDTYLIPQIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863
RESULT 8
ABG01696
ID ABG01696 standard; Protein; 2353 AA.
XX AC ABG01696;
DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #1687.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS65883.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20; SEQ ID No 32055; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2353 AA;
Query Match 99.9%; Score 9642; DB 22; Length 2353;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MDLSALRVEEVQNVINAMQKILECPICILEIKBPVSTKCDHIKFCMQLLNQKKGPSQ 60
Db 491 MDLSALRVEEVQNVINAMQKILECPICILEIKBPVSTKCDHIKFCMQLLNQKKGPSQ 550
QY 61 CPLCKNDITKRSLOESTRFSQLYVEELLKIICAFOLDGTGLEVANSYNFAKKENNSPEHLKD 120
Db 551 CPLCKNDITKRSLOESTRFSQLYVEELLKIICAFOLDGTGLEVANSYNFAKKENNSPEHLKD 610
QY 121 EVSTIQSMGYRNRAKRLLOQSEPNPSLQETSLSVQLSNLGTVRTLRTKRIQPKTSSVI 180
Db 611 EVSTIQSMGYRNRAKRLLOQSEPNPSLQETSLSVQLSNLGTVRTLRTKRIQPKTSSVI 670
QY 181 ELGSDSSEDTVNKATYCSVGDOELLQITPOGTDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 671 ELGSDSSEDTVNKATYCSVGDOELLQITPOGTDEISLDSAKKAACEFSETDVTNTEHHQ 730
QY 241 PSNNDLNTTEKRAAERHPEKYQGSSVSNLHVEPCGTWTHASSLQHENSLLLTKDRMNVE 300
Db 731 PSNNDLNTTEKRAAERHPEKYQGSSVSNLHVEPCGTWTHASSLQHENSLLLTKDRMNVE 790
QY 301 KAEEFCNKSQOPGLARSQHNWAGSKETCNDRRPTSTPEKKYVDLNADPLCERKENWKQLPC 360
Db 791 KAEEFCNKSQOPGLARSQHNWAGSKETCNDRRPTSTPEKKYVDLNADPLCERKENWKQLPC 850
QY 361 SENPRDTEVPWITLNSSIQKVNEWFSRDELGSDSDHGESESNAKVADVDLVNEVD 420
Db 851 SENPRDTEVPWITLNSSIQKVNEWFSRDELGSDSDHGESESNAKVADVDLVNEVD 910
QY 421 EYSGSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRKKASLPNLSHVTEN 480
Db 911 EYSGSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRKKASLPNLSHVTEN 970
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRRTPSGLHPEDFIKKADLAVOKTPEMINOGTNOTE 540
Db 971 LIIGAFVTEPQIIQERPLTNKLRKRRTPSGLHPEDFIKKADLAVOKTPEMINOGTNOTE 1030
QY 541 QNGQVMNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTAEPISSSISNWELELNI 600
Db 1031 QNGQVMNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTAEPISSSISNWELELNI 1090
QY 601 HNSKAPKKNLRKRSSTRHIALELVVSRNLSPNCTELOIDSCSSSEEEKKKKYNOMPV 660
Db 1091 HNSKAPKKNLRKRSSTRHIALELVVSRNLSPNCTELOIDSCSSSEEEKKKKYNOMPV 1150
QY 661 RHSRNLQMEGKEPATGAKKSNKPNQETSKRHSDTPELKLTNAPGTSFTKCSNTSELKE 720
Db 1151 RHSRNLQMEGKEPATGAKKSNKPNQETSKRHSDTPELKLTNAPGTSFTKCSNTSELKE 1210
QY 721 FVNPSPREEKEKLETVKVSNNADPKDMLSGERVLOQTORSVESSSISLVPGTDYGTQ 780
Db 1211 FVNPSPREEKEKLETVKVSNNADPKDMLSGERVLOQTORSVESSSISLVPGTDYGTQ 1270
QY 781 ESTSLLEVSTLGRKAKTEPNKCVSOCAAFENPKGLIHGCSKDNDRNDTGFKYPLGHEVNH 840
Db 1271 ESTSLLEVSTLGRKAKTEPNKCVSOCAAFENPKGLIHGCSKDNDRNDTGFKYPLGHEVNH 1330
QY 841 RETSIEMESELDAQYLQNTFKVSKRQSFALFNPNGNAEBECATFSAHSGSLKKQSPKVT 900
Db 1331 RETSIEMESELDAQYLQNTFKVSKRQSFALFNPNGNAEBECATFSAHSGSLKKQSPKVT 1390
QY 901 FECEQKEENOGKNEKNKPKVQTNITAGFPVVGOKDPVDNAKCSIKGGRFRLSSQFRG 960
Db 1391 FECEQKEENOGKNEKNKPKVQTNITAGFPVVGOKDPVDNAKCSIKGGRFRLSSQFRG 1450

Db 61 CPCKNDITKRSQESRFRQOLVEELKIIICAFQDITGLEIYANSYNPAKKNENSPHELRKD 120
QY 121 EVSTIQSMGYNRAKRLQLOSEPENPSLOETSLSVQLSNLTGTVRLTRKQRIQPKQTSVYI 180
Db 121 EVSTIQSMGYNRAKRLQLOSEPENPSLOETSLSVQLSNLTGTVRLTRKQRIQPKQTSVYI 180
QY 181 ELGSDSSEDTVNNKATYCSVGQDQELLQITPOGTRDEISLSDAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDTVNNKATYCSVGQDQELLQITPOGTRDEISLSDAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNNDLNTEKRAERHPEKXQSSVSNLHVEPCGINTHASSLOHENSULLLTCKDRNVE 300
Db 241 PSNNDLNTEKRAERHPEKXQSSVSNLHVEPCGINTHASSLOHENSULLLTCKDRNVE 300
QY 301 KAEFCNKSQOGLARSOHNWAGSKETCNDRTPTSTYKVKVDLNADPLCEKKNKQKLPC 360
Db 301 KAEFCNKSQOGLARSOHNWAGSKETCNDRTPTSTYKVKVDLNADPLCEKKNKQKLPC 360
QY 361 SENPRDTEVPWITLNSISQKVNWFSSRDELGSDSDHGESNAKAVADVLVDLNEVD 420
Db 361 SENPRDTEVPWITLNSISQKVNWFSSRDELGSDSDHGESNAKAVADVLVDLNEVD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVSSENIEDKIFGKTYRKASLPNLSHVTE 480
Db 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVSSENIEDKIFGKTYRKASLPNLSHVTE 480
QY 481 LIIGAFVTEPOIIQERPLTNKLRKRRTPSGLHPEDFIKKADLAVOKTPMINOGTNOTE 540
Db 481 LIIGAFVTEPOIIQERPLTNKLRKRRTPSGLHPEDFIKKADLAVOKTPMINOGTNOTE 540
QY 541 QNGOVNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINMELELNI 600
Db 541 QNGOVNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINMELELNI 600
QY 601 HNSKAPKNLRLKRSSTRHIALELVVSRNLSPNCTELOIDSCSSSEETKKKYNQMPV 660
Db 601 HNSKAPKNLRLKRSSTRHIALELVVSRNLSPNCTELOIDSCSSSEETKKKYNQMPV 660
QY 661 RHRNRLQMEGKEPATGAKKSNKPNQETSKRHSDTFPELKLTPNAGSFTKCSNTSELKE 720
Db 661 RHRNRLQMEGKEPATGAKKSNKPNQETSKRHSDTFPELKLTPNAGSFTKCSNTSELKE 720
QY 721 FVNPSLPREEKEKLETVKVSNAEDPKDMLSGERVLOQTERSVESSISLVPDIDYGTQ 780
Db 721 FVNPSLPREEKEKLETVKVSNAEDPKDMLSGERVLOQTERSVESSISLVPDIDYGTQ 780
QY 781 ESTSLLEVSTLGRKAKTEPNKCVSOCAAFENPKGLIHCCSKDNNDTEGFKYPLGHEVNH 840
Db 781 ESTSLLEVSTLGRKAKTEPNKCVSOCAAFENPKGLIHCCSKDNNDTEGFKYPLGHEVNH 840
QY 841 RETSIEMESELDAQYLQNTFKYSKRQSFALFSPNGNAEECATFSAHSGSLKKQSPKVT 900
Db 841 RETSIEMESELDAQYLQNTFKYSKRQSFALFSPNGNAEECATFSAHSGSLKKQSPKVT 900
QY 901 FECEQKEENOGKNESNIKPVQTNITAGFPVVGQDKPVDNAKCSIKGGSFRCLSSQFRG 960
Db 901 FECEQKEENOGKNESNIKPVQTNITAGFPVVGQDKPVDNAKCSIKGGSFRCLSSQFRG 960
QY 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVKTKCNKLNLEENFEHSMSPEREMGNENIP 1020
Db 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVKTKCNKLNLEENFEHSMSPEREMGNENIP 1020
QY 1021 STVSTISRNNIRENVFKEASSNINEVGSTNEVGSSINEIGSSDENIOAELGRNGPKL 1080
Db 1021 STVSTISRNNIRENVFKEASSNINEVGSTNEVGSSINEIGSSDENIOAELGRNGPKL 1080
QY 1081 NAMLRGLVQPEYIKOSLPGNSCKHPEIKKKQEEYEEVVQTVNTDFSPYLISDNLEQPMGSS 1140
Db 1081 NAMLRGLVQPEYIKOSLPGNSCKHPEIKKKQEEYEEVVQTVNTDFSPYLISDNLEQPMGSS 1140
QY 1141 HASQVCSETPDDLLDGEIKEDTSPAENDIKESSAVFSKVQKGLSRSPSPFTHTHLAQ 1200

Db 1141 HASQVCSETPDDLLDGEIKEDTSPAENDIKESSAVFSKVQKGLSRSPSPFTHTHLAQ 1200
QY 1201 GYRRGAKKLESSEENLSSEDEELPCFQHLHFGKVNINIPQSOTRHSHTVATECLSKNTEENL 1260
Db 1201 GYRRGAKKLESSEENLSSEDEELPCFQHLHFGKVNINIPQSOTRHSHTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDCSNOVILAKASQEHHLSEETKCSASLSFSSQCSSELEDLTANTNTQDPFLIGS 1320
Db 1261 LSLKNSLNDCSNOVILAKASQEHHLSEETKCSASLSFSSQCSSELEDLTANTNTQDPFLIGS 1320
QY 1321 SKOMRQSESQGVGLSDKELVSDDERGTGLLENNQEQSDNLSGAAAGCSESETSVSE 1380
Db 1321 SKOMRQSESQGVGLSDKELVSDDERGTGLLENNQEQSDNLSGAAAGCSESETSVSE 1380
QY 1381 DCSGLSSQSDILTTQORDTMOHNLIKLOQEMAELEAVLEQHGQSPNSYPSIISDSALE 1440
Db 1381 DCSGLSSQSDILTTQORDTMOHNLIKLOQEMAELEAVLEQHGQSPNSYPSIISDSALE 1440
QY 1441 DLNRPQSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSTSKNKEPVGVERSPSK 1500
Db 1441 DLNRPQSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSTSKNKEPVGVERSPSK 1500
QY 1501 CPSLDDRWYMHSCSGSLQNRNYPQBELIKVVDVEEQOLEESGPHDLTETSYPLODLEG 1560
Db 1501 CPSLDDRWYMHSCSGSLQNRNYPQBELIKVVDVEEQOLEESGPHDLTETSYPLODLEG 1560
QY 1561 TPYLESGISLFSDDPSDDPSDRAPESARVGNIPSTTSALKVPOLKVAESAQSPAAAHTT 1620
Db 1561 TPYLESGISLFSDDPSDDPSDRAPESARVGNIPSTTSALKVPOLKVAESAQSPAAAHTT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERNVKNRMVMVVSGLTPEEFMLVYKFARKHHTLTNLI 1680
Db 1621 DTAGYNAMESVSREKPELTASTERNVKNRMVMVVSGLTPEEFMLVYKFARKHHTLTNLI 1680
QY 1681 TEETHVVMKTDFAEFCERTLKYFLGTAGGKWVYFWTQTSIKERKMLNEHDFEVRGDV 1740
Db 1681 TEETHVVMKTDFAEFCERTLKYFLGTAGGKWVYFWTQTSIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHQQPKRARESODRKIFRGLTECCYGPFTNMPTDQLEMMVOLCGASVVKELSSFTL 1800
Db 1741 VNGRNHQQPKRARESODRKIFRGLTECCYGPFTNMPTDQLEMMVOLCGASVVKELSSFTL 1800
QY 1801 GTGVHPITVVVQPDWATEDNGFHAIGQMCEAPVVTREWVLDSDVALYQCOELDTYLIPOIPH 1860
Db 1801 GTGVHPITVVVQPDWATEDNGFHAIGQMCEAPVVTREWVLDSDVALYQCOELDTYLIPOIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863

RESULT 10
AAR81500
ID AAR81500 standard; Protein; 1863 AA.
XX
AC AAR81500;
XX
DT 02-OCT-1996 (first entry)
XX
DE BRCA1 mutant from sample sets YN98 and YN7.
XX
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 271
FT /note= "v271M"
XX
XX WO9605306-A2.
XX

PD 22-FEB-1996.
XX 11-AUG-1995; 95WO-US10202.
XX 07-JUN-1995; 95US-0483553.
PR 12-AUG-1994; 94US-0289221.
PR 02-SEP-1994; 94US-0300266.
PR 16-SEP-1994; 94US-0308104.
PR 29-NOV-1994; 94US-0348824.
PR 24-MAR-1995; 95US-0409305.
PR 07-JUN-1995; 95US-0480784.
XX (MYRI-) MYRIAD GENETICS INC.
PA (CANC-) CANCER INST.
PA (RECH-) CENT RECH DU CHUL.
XX Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
PI WPI; 1996-139702/14.
DR N-PSDB; AAT17457.
XX New nucleic acid and polypeptide for mutant or polymorphic BRCA1
PT gene - for diagnosis and therapy of human breast and ovarian cancer
PT and for diagnosing pre-disposition to these cancers
XX Claim 1; : 218pp; English.
XX AAR81483-R81497 and AAR81499-R81546 represent mutations of the protein
CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1)
CC (see AAR81481 for wild type protein). These mutations can be used as
CC immunogens for antibody production. The mutant BRCA1 genes encoding
CC these sequences have at least 1 mutation or polymorphism in comparison
CC to the wild type cDNA (see AAR17438 for wild type). By detecting a
CC germline alteration in the wild type BRCA1 gene, a predisposition for
CC breast and ovarian cancer can be diagnosed. In one method, BRCA1 mRNA
CC isolated from a tissue sample from a subject has a probe, corresponding
CC to a fragment of the cDNA encoding the wild type BRCA1 sequence (or an
CC allele-specific probe for a mutation of it), added to it. The conditions
CC allow for hybridisation of the probe to the mRNA, and any hybridisation
CC which occurs is detected. Alternatively the BRCA1 gene in the tissue
CC sample is isolated, and a shift in electrophoretic mobility of single
CC stranded DNA from the sample on a non-denaturing polyacrylamide gel
CC indicates a mutation. These methods of detection can also diagnose a
CC lesion neoplasia associated with the BRCA1 locus. The methods may be
CC used in gene therapy, protein replacement therapy and protein mimetics,
CC and may be used to screen for drugs in cancer therapy.
XX Sequence 1863 AA;
SQ Query Match 99.9%; Score 9639; DB 17; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1861; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MDLSALRVEEVQNVINAMQKILECPICLELKEPVSTKCDHIFCKFCMLLNQKKGPSQ 60
DB 1 MDLSALRVEEVQNVINAMQKILECPICLELKEPVSTKCDHIFCKFCMLLNQKKGPSQ 60
QY 61 CPLCKNDITKRSLOESTRFSQVLELLKICAFQDGTGLEVANSYNFAKKNNSPEHLKD 120
DB 61 CPLCKNDITKRSLOESTRFSQVLELLKICAFQDGTGLEVANSYNFAKKNNSPEHLKD 120
QY 121 EYSIIQSMGYRNRAKLQSEPNFSLOETSLVSQSLNLTGTVTLTKRQIQPKTSYVI 180
DB 121 EYSIIQSMGYRNRAKLQSEPNFSLOETSLVSQSLNLTGTVTLTKRQIQPKTSYVI 180
QY 181 ELGSDSSEDTVNKATYCSVGDOELQITPQGTTRDEISLSAKKAACEFSETDVTNTEHHQ 240
DB 181 ELGSDSSEDTVNKATYCSVGDOELQITPQGTTRDEISLSAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNNDLNTTEKRAAERHPKYGSSVSNLHVEPCGTNTNTHASSLQHENSLLLTDRMNV 300
DB 241 PSNNDLNTTEKRAAERHPKYGSSVSNLHVEPCGTNTNTHASSLQHENSLLLTDRMNV 300

QY 301 KAEFCNKSQGLARSOHNRWAGSKETCNDRTPTSTKVKVDLNADPLCERKWNKOKLPC 360
DB 301 KAEFCNKSQGLARSOHNRWAGSKETCNDRTPTSTKVKVDLNADPLCERKWNKOKLPC 360
QY 361 SENPRDTEVPWITLNSSTQKVNEWFSRSDLLGSDSDHGESESNKAVADVLDVLYNEVD 420
DB 361 SENPRDTEVPWITLNSSTQKVNEWFSRSDLLGSDSDHGESESNKAVADVLDVLYNEVD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRKKASLNLSHVTEEN 480
DB 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRKKASLNLSHVTEEN 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPDEFKKAADLAVOKTPEMINOQNTQTE 540
DB 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPDEFKKAADLAVOKTPEMINOQNTQTE 540
QY 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINMELELNI 600
DB 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINMELELNI 600
QY 601 HNSKAPKNRLRRKSTRHIIHALELVVSRNLSPPNCTELQIDSCSSSEBIKKKYNQMPV 660
DB 601 HNSKAPKNRLRRKSTRHIIHALELVVSRNLSPPNCTELQIDSCSSSEBIKKKYNQMPV 660
QY 661 RHRNLQMEGKEPATGAKKSNKPNEQTSKRHSDTEPELKLNTNAPGSKTCSNTSELKE 720
DB 661 RHRNLQMEGKEPATGAKKSNKPNEQTSKRHSDTEPELKLNTNAPGSKTCSNTSELKE 720
QY 721 FVNPSLPREEKEELKLVKVSNNNAEDPKDMLSGERVLTQTERSVESSISLVPGTDYGTQ 780
DB 721 FVNPSLPREEKEELKLVKVSNNNAEDPKDMLSGERVLTQTERSVESSISLVPGTDYGTQ 780
QY 781 EISILLEVSTLIGKAKTEPNKCVSQCAAFENPKGLIHGSKDNRNDETEGFKYPLGHEVNH 840
DB 781 EISILLEVSTLIGKAKTEPNKCVSQCAAFENPKGLIHGSKDNRNDETEGFKYPLGHEVNH 840
QY 841 RETSTEMEESELDQYQLONTFKVSKRQSFALFSPNPGNAEEECATFSAHSGSLKKQSPKVT 900
DB 841 RETSTEMEESELDQYQLONTFKVSKRQSFALFSPNPGNAEEECATFSAHSGSLKKQSPKVT 900
QY 901 FECQKEENQGNESNIKPVTQNTITAGFPVVGQKDPVDNAKCSIKGSRFCLSSQFRG 960
DB 901 FECQKEENQGNESNIKPVTQNTITAGFPVVGQKDPVDNAKCSIKGSRFCLSSQFRG 960
QY 961 NETGLITPNKHGLQNPYRIPPLPIKSFVTKCKKNLLEENFEHSHMSPEREMGNENIP 1020
DB 961 NETGLITPNKHGLQNPYRIPPLPIKSFVTKCKKNLLEENFEHSHMSPEREMGNENIP 1020
QY 1021 STVSTISRNNITRENVFKEASSNINEVGSSNINEVGSSINEIGSSDENTQAEELGRNRPKL 1080
DB 1021 STVSTISRNNITRENVFKEASSNINEVGSSNINEVGSSINEIGSSDENTQAEELGRNRPKL 1080
QY 1081 NAMLRGLVQPEVYKQSLPGSNCKHPEIKKOEYEVWQTVNTDFSPYLLISDLEQPMGSS 1140
DB 1081 NAMLRGLVQPEVYKQSLPGSNCKHPEIKKOEYEVWQTVNTDFSPYLLISDLEQPMGSS 1140
QY 1141 HASQVCSPTDDLLDDGEIKEDTSFAENDIKESSAVFSKSVQKGLSRSPSFTHHLAQ 1200
DB 1141 HASQVCSPTDDLLDDGEIKEDTSFAENDIKESSAVFSKSVQKGLSRSPSFTHHLAQ 1200
QY 1201 GYRGAKKLESSEENLSEDEELPCFQHLKFKVNNIPFSQTRHSTVATECLSKNTEENL 1260
DB 1201 GYRGAKKLESSEENLSEDEELPCFQHLKFKVNNIPFSQTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDSCNQVILAKASQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS 1320
DB 1261 LSLKNSLNDSCNQVILAKASQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS 1320
QY 1321 SKQMRHOSQGVGLSKDELVDDEERTGLEENNOESQSDNSLGEASGCESETSYSE 1380
DB 1321 SKQMRHOSQGVGLSKDELVDDEERTGLEENNOESQSDNSLGEASGCESETSYSE 1380
QY 1381 DCSGLSSOSDILTTTQQRDTMOHNLTKLOQEMAEAVLEQHSQSPSNSYPSTISDSSALE 1440

Db 1381 DCSGLSQSDILTOQDTHQHNLIKLQEMAELEAVLEHGQSPNSYPSIISDSSALE 1440
QY 1441 DLNRPQSTSEKAVLTSQKSSSEYPISONPEGLSADKFEVSADSTSKNKEPGVERSPSK 1500
Db 1441 DLNRPQSTSEKAVLTSQKSSSEYPISONPEGLSADKFEVSADSTSKNKEPGVERSPSK 1500
QY 1501 CPSLDRWYMHSCSGSLQNRNYPQBEELIKVVDVEEQQLLEESGPHDLTETSYLPRQDLEG 1560
Db 1501 CPSLDRWYMHSCSGSLQNRNYPQBEELIKVVDVEEQQLLEESGPHDLTETSYLPRQDLEG 1560
QY 1561 TPYLESGISLFSDDPSDDPSDDRAPSARVGNTPSTSAKVPQLKVAESAQSPAHAHTT 1620
Db 1561 TPYLESGISLFSDDPSDDPSDDRAPSARVGNTPSTSAKVPQLKVAESAQSPAHAHTT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEEFMLVYKFARKKHITITNLI 1680
Db 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEEFMLVYKFARKKHITITNLI 1680
QY 1681 TEETHVWVKTDAEFCERTKYFLGIAGGKWVYVFWVTQTSIKERKMLNEHDFEVRGDV 1740
Db 1681 TEETHVWVKTDAEFCERTKYFLGIAGGKWVYVFWVTQTSIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRHHQPKRARESQRKIFRGLGIECCYGPFTNMPTDQLEWVQLCGASVVKELSSFTL 1800
Db 1741 VNGRHHQPKRARESQRKIFRGLGIECCYGPFTNMPTDQLEWVQLCGASVVKELSSFTL 1800
QY 1801 GTGVHPIVVVQPDAMWEDNGFHAIGOMCEAPVVTREWWLDSVALYOCQELDTYLIPQIPH 1860
Db 1801 GTGVHPIVVVQPDAMWEDNGFHAIGOMCEAPVVTREWWLDSVALYOCQELDTYLIPQIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863

RESULT 11
AAR81522
ID AAR81522 standard; Protein; 1863 AA.
AC AAR81522;
XX
DT 02-OCT-1996 (first entry)
XX
DE BRCA1 mutant from sample set MSK12871.
XX
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
XX
OS Homo sapiens.
XX
FH Key
FT Misc-difference 1219 /note= "E1219D"
XX
PN WO9605306-A2.
XX
PD 22-FEB-1996.
XX
PF 11-AUG-1995; 95WO-US10202.
XX
PR 07-JUN-1995; 95US-0483553.
PR 12-AUG-1994; 94US-0289221.
PR 02-SEP-1994; 94US-0300266.
PR 16-SEP-1994; 94US-0308104.
PR 29-NOV-1994; 94US-0348824.
PR 24-MAR-1995; 95US-0409305.
PR 07-JUN-1995; 95US-0480784.
XX
PA (MYRI-) MYRIAD GENETICS INC.
PA (CANC-) CANCER INST.
PA (RECH-) CENT RECH DU CHUL.

XX Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
XX WPI: 1996-139702/14.
DR N-PSDB; AAT17479.
XX
PT New nucleic acid and polypeptide for mutant or polymorphic BRCA1
PT gene - for diagnosis and therapy of human breast and ovarian cancer
PT and for diagnosing pre-disposition to these cancers
XX
PS Claim 1: 218pp; English.
XX
CC AAR81483-R81497 and AAR81499-R81546 represent mutations of the protein
CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1)
CC (see AAR81481 for wild type protein). These mutations can be used as
CC immunogens for antibody production. The mutant BRCA1 genes encoding
CC these sequences have at least 1 mutation or polymorphism in comparison
CC to the wild type cDNA (see AAT17438 for wild type). By detecting a
CC germline alteration in the wild type BRCA1 gene, a predisposition for
CC breast and ovarian cancer can be diagnosed. In one method, BRCA1 mRNA
CC isolated from a tissue sample from a subject has a probe, corresponding
CC to a fragment of the cDNA encoding the wild type BRCA1 sequence (or an
CC allele-specific probe for a mutation of it), added to it. The conditions
CC allow for hybridisation of the probe to the mRNA, and any hybridisation
CC which occurs is detected. Alternatively the BRCA1 gene in the tissue
CC sample is isolated, and a shift in electrophoretic mobility of single
CC stranded DNA from the sample on a non-denaturing polyacrylamide gel
CC indicates a mutation. These methods of detection can also diagnose a
CC lesion neoplasia associated with the BRCA1 locus. The methods may be
CC used in gene therapy, protein replacement therapy and protein mimetics,
CC and may be used to screen for drugs in cancer therapy.
XX
XX Sequence 1863 AA;
SQ
Query Match 99.9%; Score 9639; DB 17; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1861; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MDLSALRVEEVQNVINAMOKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNKKGPQSQ 60
Db 1 MDLSALRVEEVQNVINAMOKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNKKGPQSQ 60
QY 61 CPLCKNDITKRSLOESTRFSQVVEELLKIICAFQDGTGLYANSYNAFKAENNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRFSQVVEELLKIICAFQDGTGLYANSYNAFKAENNSPEHLKD 120
QY 121 EYSIIOSMGYRNRAKRLLOSEPENPSLQFTSLSVQSLNLTGTVRTLTKRIOPQKTSVYI 180
Db 121 EYSIIOSMGYRNRAKRLLOSEPENPSLQFTSLSVQSLNLTGTVRTLTKRIOPQKTSVYI 180
QY 181 ELGSDSSEDTVNKATYCSVGDOFELLOITPQGTREISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDTVNKATYCSVGDOFELLOITPQGTREISLDSAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNNDLNTTEKRAAERHPKYQSSVSNLHVEPCGNTNTHASSLQHENSILLTKDRMNYE 300
Db 241 PSNNDLNTTEKRAAERHPKYQSSVSNLHVEPCGNTNTHASSLQHENSILLTKDRMNYE 300
QY 301 KAEFCNKSKQPLARSOHNRWAGSKETCNDRRTPSTPEKKVDLNLNADPLCERKENNKOKLPC 360
Db 301 KAEFCNKSKQPLARSOHNRWAGSKETCNDRRTPSTPEKKVDLNLNADPLCERKENNKOKLPC 360
QY 361 SENPRDTEVPWITLNSSTOKYNEWFSRDELLGSDSDSHGSESNKAVADVLDVLENYD 420
Db 361 SENPRDTEVPWITLNSSTOKYNEWFSRDELLGSDSDSHGSESNKAVADVLDVLENYD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRKKASLPNLSHVTEN 480
Db 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRKKASLPNLSHVTEN 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPMINSQGTNOTE 540
Db 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPMINSQGTNOTE 540

QY 541 QNGQVNNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNMELELNI 600
DB 541 QNGQVNNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNMELELNI 600
QY 601 HNSKAPKNRLRRKSTRIHIALELWVRNLSPPNCTELQIDSCSSSEIEIKKKYQNPV 660
DB 601 HNSKAPKNRLRRKSTRIHIALELWVRNLSPPNCTELQIDSCSSSEIEIKKKYQNPV 660
QY 661 RHSRNLQMEGKEPATGAKSKNPNQOTSKRHDSPTFPELKLITNAPGSFTKCSNTSELKE 720
DB 661 RHSRNLQMEGKEPATGAKSKNPNQOTSKRHDSPTFPELKLITNAPGSFTKCSNTSELKE 720
QY 721 FVNPSLPREKEKELETYKVSNNADDPKDLMLSGSERVLQTERSVESSISLVPDXTGQT 780
DB 721 FVNPSLPREKEKELETYKVSNNADDPKDLMLSGSERVLQTERSVESSISLVPDXTGQT 780
QY 781 ESIISLLEVTGLKAKTEPNKVCQCAAFENPKGLIHGCKDNRNDTEGFKYPLGHEVNH 840
DB 781 ESIISLLEVTGLKAKTEPNKVCQCAAFENPKGLIHGCKDNRNDTEGFKYPLGHEVNH 840
QY 841 RETSIEMESELDAOYLONTFKVSKRQSFALFSPNPGNAEEBCATFSAHSGSLKKQSPKVT 900
DB 841 RETSIEMESELDAOYLONTFKVSKRQSFALFSPNPGNAEEBCATFSAHSGSLKKQSPKVT 900
QY 901 FECEQKEENOGKNESNKPQVTNITAGFPVVGQDKPVDNAKCSIKGSRFCLSSQFRG 960
DB 901 FECEQKEENOGKNESNKPQVTNITAGFPVVGQDKPVDNAKCSIKGSRFCLSSQFRG 960
QY 961 NETGLITPNKHGLLQNPRIPLPIKSFVTKCKKKNLLENFEBHSGSPEREMGNENIP 1020
DB 961 NETGLITPNKHGLLQNPRIPLPIKSFVTKCKKKNLLENFEBHSGSPEREMGNENIP 1020
QY 1021 SVVSTISRNNIRENVFKKASSNINEVGSSSTNEIGSSDENIQAEELGRNRGPKL 1080
DB 1021 SVVSTISRNNIRENVFKKASSNINEVGSSSTNEIGSSDENIQAEELGRNRGPKL 1080
QY 1081 NAMRLGLVQLPEVYKQSLPGSNCKHPEIKQIEYEVVQTVNTDFSPYLISDNLEQPMGSS 1140
DB 1081 NAMRLGLVQLPEVYKQSLPGSNCKHPEIKQIEYEVVQTVNTDFSPYLISDNLEQPMGSS 1140
QY 1141 HASQVCSETPDDLDDGEIKEDTSPAENDIKESSAVFSKSVQKGLSRSPSPFTHLQAQ 1200
DB 1141 HASQVCSETPDDLDDGEIKEDTSPAENDIKESSAVFSKSVQKGLSRSPSPFTHLQAQ 1200
QY 1201 GYRRGAKKLESSEENLSDDEELPCFQHLFLGKYNNTIPSQSTRSTVATECLSKNTEENL 1260
DB 1201 GYRRGAKKLESSEENLSDDEELPCFQHLFLGKYNNTIPSQSTRSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDCSNOVILAKASQEHHLSEETKCSASLFSQCSELEDLTANTNTQDPFLIGS 1320
DB 1261 LSLKNSLNDCSNOVILAKASQEHHLSEETKCSASLFSQCSELEDLTANTNTQDPFLIGS 1320
QY 1321 SKQMRHQSQGVGLSDKLEYSDDERGTGLEENNQBESQMSNLGAAAGCESETSVSE 1380
DB 1321 SKQMRHQSQGVGLSDKLEYSDDERGTGLEENNQBESQMSNLGAAAGCESETSVSE 1380
QY 1381 DCSGLSSOSDILTTQOORTMOHNLKLOEMAELVLEQHSQSPSNYSPIIISDSSALE 1440
DB 1381 DCSGLSSOSDILTTQOORTMOHNLKLOEMAELVLEQHSQSPSNYSPIIISDSSALE 1440
QY 1441 DLURNPEQSTSEKAVLTQSSEYPISQNPEGLSADKFEVSADSSSTSKNKEFGVERSSPSK 1500
DB 1441 DLURNPEQSTSEKAVLTQSSEYPISQNPEGLSADKFEVSADSSSTSKNKEFGVERSSPSK 1500
QY 1501 CPSLDRWYMHSCSGSLQNRNYPQSEELIKVVDVEQOLESRGPDLTETYSYLPQDLEG 1560
DB 1501 CPSLDRWYMHSCSGSLQNRNYPQSEELIKVVDVEQOLESRGPDLTETYSYLPQDLEG 1560
QY 1561 TPYLESGLSLFSDDPSPEDPAPESARVGNIPSSTSALKVPOLKVAESAQSPAAAHTT 1620
DB 1561 TPYLESGLSLFSDDPSPEDPAPESARVGNIPSSTSALKVPOLKVAESAQSPAAAHTT 1620

QY 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPPEEFMLVYKFARKHHITLTNLI 1680
DB 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPPEEFMLVYKFARKHHITLTNLI 1680
QY 1681 TEETHVVMKTDADFVCERTLKYPFLGIAGGKVVYSYFWVTQSIIKERKMLNEHDFEVRGDV 1740
DB 1681 TEETHVVMKTDADFVCERTLKYPFLGIAGGKVVYSYFWVTQSIIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHOGPKRARESODRKIFRGLIEICCYGPTTNMPTDQLEMMVOLCGASVVKELSSFTL 1800
DB 1741 VNGRNHOGPKRARESODRKIFRGLIEICCYGPTTNMPTDQLEMMVOLCGASVVKELSSFTL 1800
QY 1801 GTGVHPVVVQPDAWTEDNGFHAIGQMCEAPVVTREWVLDSVALYQCOELDTYLIPQIPH 1860
DB 1801 GTGVHPVVVQPDAWTEDNGFHAIGQMCEAPVVTREWVLDSVALYQCOELDTYLIPQIPH 1860
QY 1861 SHY 1863
DB 1861 SHY 1863
RESULT 12
AAR81536
ID AAR81536 standard; Protein; 1863 AA.
XX AAR81536;
XX AC
XX DT 02-OCT-1996 (first entry)
XX DE
XX KW
XX KW
XX KW
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Misc-difference 1183
XX FT /note= "K1183R"
XX PN WO9605306-A2.
XX PD 22-FEB-1996.
XX PF 11-AUG-1995; 95WO-US10202.
XX PR 07-JUN-1995; 95US-0483553.
PR 12-AUG-1994; 94US-0289221.
PR 02-SEP-1994; 94US-0300266.
PR 16-SEP-1994; 94US-0308104.
PR 29-NOV-1994; 94US-0348824.
PR 24-MAR-1995; 95US-0409305.
PR 07-JUN-1995; 95US-0480784.
XX (MYRI-) MYRIAD GENETICS INC.
PA (CANC-) CANCER INST.
PA (RECH-) CENT RECH DU CHUL.
XX Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
XX WPI; 1996-139702/14.
DR N-PSDB; AAT17494.
XX
PT New nucleic acid and polypeptide for mutant or polymorphic BRCA1
PT gene - for diagnosis and therapy of human breast and ovarian cancer
PT and for diagnosing pre-disposition to these cancers
XX
XX Claim 1; : 218pp; English.
PS
CC AAR81483-R81497 and AAR81499-R81546 represent mutations of the protein
CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1)
CC (see AAR81481 for wild type protein). These mutations can be used as

CC immunogens for antibody production. The mutant BRCA1 genes encoding
CC these sequences have at least 1 mutation or polymorphism in comparison
CC to the wild type cDNA (see AAT17438 for wild type). By detecting a
CC germline alteration in the wild type BRCA1 gene, a predisposition for
CC breast and ovarian cancer can be diagnosed. In one method, BRCA1 mRNA
CC isolated from a tissue sample from a subject has a probe, corresponding
CC to a fragment of the cDNA encoding the wild type BRCA1 sequence (or an
CC allele-specific probe for a mutation of it), added to it. The conditions
CC allow for hybridisation of the probe to the mRNA, and any hybridisation
CC which occurs is detected. Alternatively the BRCA1 gene in the tissue
CC sample is isolated, and a shift in electrophoretic mobility of single
CC stranded DNA from the sample on a non-denaturing polyacrylamide gel
CC indicates a mutation. These methods of detection can also diagnose a
CC lesion neoplasia associated with the BRCA1 locus. The methods may be
CC used in gene therapy, protein replacement therapy and protein mimetics,
CC and may be used to screen for drugs in cancer therapy.

XX
SQ Sequence 1863 AA;

Query Match 99.9%; Score 9639; DB 17; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1861; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDLSALRVVEQVNVINAMQILECPICILELIKPEVSTKCDHIFCKFCMLKLLNKKGPSQ 60
DB 1 MDLSALRVVEQVNVINAMQILECPICILELIKPEVSTKCDHIFCKFCMLKLLNKKGPSQ 60

QY 61 CPICKNDITKRSLOESTRFQVVEELKIIICAFOLDGLEYANSYNFAKKENNSPEHLKD 120
DB 61 CPICKNDITKRSLOESTRFQVVEELKIIICAFOLDGLEYANSYNFAKKENNSPEHLKD 120

QY 121 EVSIQSMGYRNRAKRLQLOSEPENPSLOETSLSVQLSNLGTVRTLRKQIQPKTSVYI 180
DB 121 EVSIQSMGYRNRAKRLQLOSEPENPSLOETSLSVQLSNLGTVRTLRKQIQPKTSVYI 180

QY 181 ELGSDSSEDIVNKATYCSVGQDQLLQITPGQTRDEISLDSAKKAACSEFSDVTNTSEHQ 240
DB 181 ELGSDSSEDIVNKATYCSVGQDQLLQITPGQTRDEISLDSAKKAACSEFSDVTNTSEHQ 240

QY 241 PSNNDLNTTEKRAAERPEKYQSSVSNLHVPCGNTTHASSLQHENSLLLTKDRMNV 300
DB 241 PSNNDLNTTEKRAAERPEKYQSSVSNLHVPCGNTTHASSLQHENSLLLTKDRMNV 300

QY 301 KAECFNKSKOPGLARSQHNWAGSKETCNDRRTPSTPEKKYVDLNADPLCKERKWNKQLPC 360
DB 301 KAECFNKSKOPGLARSQHNWAGSKETCNDRRTPSTPEKKYVDLNADPLCKERKWNKQLPC 360

QY 361 SENPRDTEVPWITLNSSIQKVNEWFSRDELGSDSDSHGSESNAKVADVLVDLNEVD 420
DB 361 SENPRDTEVPWITLNSSIQKVNEWFSRDELGSDSDSHGSESNAKVADVLVDLNEVD 420

QY 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRKKASLPNLSHVTE 480
DB 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRKKASLPNLSHVTE 480

QY 481 LIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPEDFIKKADLAVQKTPEMINQGTNOTE 540
DB 481 LIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPEDFIKKADLAVQKTPEMINQGTNOTE 540

QY 541 QNGOVNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNWELENI 600
DB 541 QNGOVNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNWELENI 600

QY 601 HNSKAPKKNLRKRSRTHALELVYSRNLSPNCPNTELOIDSCSSSEEEKKKKYNOMPV 660
DB 601 HNSKAPKKNLRKRSRTHALELVYSRNLSPNCPNTELOIDSCSSSEEEKKKKYNOMPV 660

QY 661 RHSNRLQMECKEPATGAKSKSNPNEQTSKRHSDTFPELKLTNAPGSFTKCSNTSELKE 720
DB 661 RHSNRLQMECKEPATGAKSKSNPNEQTSKRHSDTFPELKLTNAPGSFTKCSNTSELKE 720

QY 721 FVNPSPREEKEKLETVKVSNNAEADPKDMLSGERVLOTERSVESSSISLVPGDTGYGTQ 780
DB 721 FVNPSPREEKEKLETVKVSNNAEADPKDMLSGERVLOTERSVESSSISLVPGDTGYGTQ 780

DB 721 FVNPSPREEKEKLETVKVSNNAEADPKDMLSGERVLOTERSVESSSISLVPGDTGYGTQ 780

QY 781 ESTSLLEVSTLGRAKTEPNKCVSOCAAFENPKGLIHGCSKDNRRNDTEGKYPLGHEVNH 840
DB 781 ESTSLLEVSTLGRAKTEPNKCVSOCAAFENPKGLIHGCSKDNRRNDTEGKYPLGHEVNH 840

QY 841 RETSIEMESELDAQYLQNTFKVSKRQSFALFNSPGNAEEECATFSAHSGSLKKQSPKVT 900
DB 841 RETSIEMESELDAQYLQNTFKVSKRQSFALFNSPGNAEEECATFSAHSGSLKKQSPKVT 900

QY 901 FECEQKEENOGKKNESNKPQVTNITAGFPVVGOKDPVDNAKCSIKGSRFCLSSQFRG 960
DB 901 FECEQKEENOGKKNESNKPQVTNITAGFPVVGOKDPVDNAKCSIKGSRFCLSSQFRG 960

QY 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVKTKCKNLLLEENFEHSMSPEREMGNENIP 1020
DB 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVKTKCKNLLLEENFEHSMSPEREMGNENIP 1020

QY 1021 STVSTISRNNIRENVFKAEASSNNINEVGSSTNEVGSSEINIGSSDENIQIAELGRNRPKL 1080
DB 1021 STVSTISRNNIRENVFKAEASSNNINEVGSSTNEVGSSEINIGSSDENIQIAELGRNRPKL 1080

QY 1081 NAMRLGVLOPEVYKQSLPGSNCKHPEIKKOEYEEVVQTVNTDFSPYLIISDNLEQPMGSS 1140
DB 1081 NAMRLGVLOPEVYKQSLPGSNCKHPEIKKOEYEEVVQTVNTDFSPYLIISDNLEQPMGSS 1140

QY 1141 HASQVCSETPDDLDDGEIKEDTSFAENDIKESAVFSKVGKELSRSPPTHTHLAQ 1200
DB 1141 HASQVCSETPDDLDDGEIKEDTSFAENDIKESAVFSKVGKELSRSPPTHTHLAQ 1200

QY 1201 GYRRGAKKLESSEENLSSEDEELPCFOHLLFGKVNINPISOSTRHSTVATECLSKNTEENL 1260
DB 1201 GYRRGAKKLESSEENLSSEDEELPCFOHLLFGKVNINPISOSTRHSTVATECLSKNTEENL 1260

QY 1261 LSLKNSLDCSNQVILAKASQEHHLSEETKCSASLFSQCSSELEDLTANTQDPFLIGS 1320
DB 1261 LSLKNSLDCSNQVILAKASQEHHLSEETKCSASLFSQCSSELEDLTANTQDPFLIGS 1320

QY 1321 SKOMRHQSESGVGLSDKELVSDDEERGTLGLENNEQEQSMDSNLGEAASGCSETSVSE 1380
DB 1321 SKOMRHQSESGVGLSDKELVSDDEERGTLGLENNEQEQSMDSNLGEAASGCSETSVSE 1380

QY 1381 DCSGLSSQSDILTTQQRDRTMOHNLIKLOQEMAELEAVLEQHGQSPNSYPSIISDSSALE 1440
DB 1381 DCSGLSSQSDILTTQQRDRTMOHNLIKLOQEMAELEAVLEQHGQSPNSYPSIISDSSALE 1440

QY 1441 DLRNPEOSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSTSKNKEPGEVERSPSK 1500
DB 1441 DLRNPEOSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSTSKNKEPGEVERSPSK 1500

QY 1501 CPSLDDRWYMHSCSGSLQNRNYPQOELIKVVDVEEQOLEESGPHDLTETSYLPRDLEG 1560
DB 1501 CPSLDDRWYMHSCSGSLQNRNYPQOELIKVVDVEEQOLEESGPHDLTETSYLPRDLEG 1560

QY 1561 TPYLESGISLFSDDPSDESDRAPESARVGNIPSTTSALKVPQLKVAESAQSPAAAHTT 1620
DB 1561 TPYLESGISLFSDDPSDESDRAPESARVGNIPSTTSALKVPQLKVAESAQSPAAAHTT 1620

QY 1621 DTAGYNAMESVSREKPELTAFTERVNRKSMVMVSGLTPEEFMLVYKFARKHHTLTNLI 1680
DB 1621 DTAGYNAMESVSREKPELTAFTERVNRKSMVMVSGLTPEEFMLVYKFARKHHTLTNLI 1680

QY 1681 TEETHVMKTDFAEVCERTLKVFGLIAGCKWVYVFWVTQSTIKERKMLNEHDFEVRGDV 1740
DB 1681 TEETHVMKTDFAEVCERTLKVFGLIAGCKWVYVFWVTQSTIKERKMLNEHDFEVRGDV 1740

QY 1741 VNGRNHQGPKRARESQRDKIFRGLETCCYGPFTNMPTDQLEWMVQLCGASVVKSELSSFTL 1800
DB 1741 VNGRNHQGPKRARESQRDKIFRGLETCCYGPFTNMPTDQLEWMVQLCGASVVKSELSSFTL 1800

QY 1801 GTGVHPITVVVQPDWATEDNGFHAIGQMCEAPVVTREWLDSVALYQCOELDTYLIPOIPH 1860
DB 1801 GTGVHPITVVVQPDWATEDNGFHAIGQMCEAPVVTREWLDSVALYQCOELDTYLIPOIPH 1860

QY 1861 SHY 1863
III
Db 1861 SHY 1863

RESULT 13
AAR81540
ID AAR81540 standard; Protein; 1863 AA.
XX AAR81540;
XX
XX 02-OCT-1996 (first entry)
XX
XX BRCA1 mutant from PM20.
XX
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT Misc-difference 1040 /note= "S1040N"
FT
XX
XX W096050306-A2.
XX
XX
XX 22-FEB-1996.
XX
XX 11-AUG-1995; 95WO-US10202.
XX
XX 07-JUN-1995; 95US-0483553.
XX 12-AUG-1994; 94US-0289221.
XX 02-SEP-1994; 94US-0300266.
XX 16-SEP-1994; 94US-0308104.
XX 29-NOV-1994; 94US-0348824.
XX 24-MAR-1995; 95US-0409305.
XX 07-JUN-1995; 95US-0480784.
XX
PA (MYRI-) MYRIAD GENETICS INC.
PA (CANC-) CANCER INST.
PA (RECH-) CENT RECH DU CHUL.
XX
XX Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
WPI: 1996-139702/14.
DR N-PSDB; AAT17502.
XX
XX New nucleic acid and polypeptide for mutant or polymorphic BRCA1
PT gene - for diagnosis and therapy of human breast and ovarian cancer
PT and for diagnosing pre-disposition to these cancers
PT
XX
XX Claim 1; ; 218pp; English.

AAR81483-R81497 and AAR81499-R81546 represent mutations of the protein encoded by the human breast and ovarian cancer predisposing gene (BRCA1) (see AAR81481 for wild type protein). These mutations can be used as immunogens for antibody production. The mutant BRCA1 genes encoding these sequences have at least 1 mutation or polymorphism in comparison to the wild type cDNA (see AAT17438 for wild type). By detecting a germline alteration in the wild type BRCA1 gene, a predisposition for breast and ovarian cancer can be diagnosed. In one method, BRCA1 mRNA isolated from a tissue sample from a subject has a probe, corresponding to a fragment of the cDNA encoding the wild type BRCA1 sequence (or an allele-specific probe for a mutation of it), added to it. The conditions allow for hybridisation of the probe to the mRNA, and any hybridisation which occurs is detected. Alternatively the BRCA1 gene in the tissue sample is isolated, and a shift in electrophoretic mobility of single stranded DNA from the sample on a non-denaturing polyacrylamide gel indicates a mutation. These methods of detection can also diagnose a lesion neoplasia associated with the BRCA1 locus. The methods may be used in gene therapy, protein replacement therapy and protein mimetics,

CC and may be used to screen for drugs in cancer therapy.
XX
SQ Sequence 1863 AA;
Query Match 99.9%; Score 9639; DB 17; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1861; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MDLSALRVVEVQVINAMOKIIECPICLLEIKPEVSTKCDHIFCKFCMLKLLNKKGPSQ 60
Db 1 MDLSALRVVEVQVINAMOKIIECPICLLEIKPEVSTKCDHIFCKFCMLKLLNKKGPSQ 60
QY 61 CPLCKNDITKRSLOESTRFSQVLELLKTIICAFQDITGLEYANSYNFAKKNNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRFSQVLELLKTIICAFQDITGLEYANSYNFAKKNNSPEHLKD 120
QY 121 EYSIIQSMGYNRNKRLLQSEPNPSLOETSLVQSLNLTGVTTLTKRIQIQKTSVYI 180
Db 121 EYSIIQSMGYNRNKRLLQSEPNPSLOETSLVQSLNLTGVTTLTKRIQIQKTSVYI 180
QY 181 ELGSDSSEDTYNKATYCSVGDOELLQITPQGTREISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDTYNKATYCSVGDOELLQITPQGTREISLDSAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNDLNTTEKRAERHPEKYQGSVSNLHVPCGTNTHASSLOHENSLLLTKDRMNVE 300
Db 241 PSNDLNTTEKRAERHPEKYQGSVSNLHVPCGTNTHASSLOHENSLLLTKDRMNVE 300
QY 301 KAEFCNKSQKGLARSQHNWAGSKETCNDRTPTSTKVKVDLNADPLCERKEWNKKLP 360
Db 301 KAEFCNKSQKGLARSQHNWAGSKETCNDRTPTSTKVKVDLNADPLCERKEWNKKLP 360
QY 361 SENPRDTEVPWITLNSSTQKWNFSDLLSGDSDHGESENAKVADVDLVNEVD 420
Db 361 SENPRDTEVPWITLNSSTQKWNFSDLLSGDSDHGESENAKVADVDLVNEVD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKASLPNLSHVTEN 480
Db 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKASLPNLSHVTEN 480
QY 481 LIIGAFVTEPQIIQERPLTNLKKRRRPTSGLHPDEFKKADLAVOKTPEMINQGTNQE 540
Db 481 LIIGAFVTEPQIIQERPLTNLKKRRRPTSGLHPDEFKKADLAVOKTPEMINQGTNQE 540
QY 541 QNGQVMNITNSGHNKTKGDSIQNEKNPNPIESLKESEAFKTKAEPISISSINMELELNI 600
Db 541 QNGQVMNITNSGHNKTKGDSIQNEKNPNPIESLKESEAFKTKAEPISISSINMELELNI 600
QY 601 HNSKAPKNRLRRKSSRHHIHALELVSRNLSPPNCTELQIDCSSSEIEIKKKYQMPV 660
Db 601 HNSKAPKNRLRRKSSRHHIHALELVSRNLSPPNCTELQIDCSSSEIEIKKKYQMPV 660
QY 661 RHRNLQJMEGKEPATGAKSNKNEQTSKRHSDSTFPPELKLITNAPGSKTSNTSELKE 720
Db 661 RHRNLQJMEGKEPATGAKSNKNEQTSKRHSDSTFPPELKLITNAPGSKTSNTSELKE 720
QY 721 FVNPSLPREEKEELKLVKVSNNADPKDMLSGERVLTQTSRVSSSISLVPGTDYGTQ 780
Db 721 FVNPSLPREEKEELKLVKVSNNADPKDMLSGERVLTQTSRVSSSISLVPGTDYGTQ 780
QY 781 ESIISLLEVTSLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNNDTEGFKYPLGHEVNH 840
Db 781 ESIISLLEVTSLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNNDTEGFKYPLGHEVNH 840
QY 841 RETSIEMESELDAQYLQNTFKVSRQSFALFSPNCAEEECATFSAHSGSLKKQSPKVT 900
Db 841 RETSIEMESELDAQYLQNTFKVSRQSFALFSPNCAEEECATFSAHSGSLKKQSPKVT 900
QY 901 FECQKEENQCKNESNIKPVQTVNITAGFPVVGQDKPVDNAKCSIKGSGRCLSSQFRG 960
Db 901 FECQKEENQCKNESNIKPVQTVNITAGFPVVGQDKPVDNAKCSIKGSGRCLSSQFRG 960
QY 961 NETGLITPNKHGLLQNPYRIPPLFPKISFVKTKCKKNLLEENFEHSMSPERMGNIPI 1020

```

Db 961  |||||NETGLITPNKHG|||LQNPYRIPPLFPKSFVKTKCKNLLN|EENFEHSP|ERDMGN|ENIP 1020
QY 1021  |||||STVSTIRNNIRENVFKEASSNINEVGSTNEVGSINEIGSSDENIOAELGRN|GPKL 1080
Db 1021  |||||STVSTIRNNIRENVFKEASSNINEVGSTNEVGSINEIGSSDENIOAELGRN|GPKL 1080
QY 1081  |||||NAMLRLGVQLQPEYKQSLPGSNC|KHPKQYEEVYVQVNTDFSPYLLSDNLEQPMGSS 1140
Db 1081  |||||NAMLRLGVQLQPEYKQSLPGSNC|KHPKQYEEVYVQVNTDFSPYLLSDNLEQPMGSS 1140
QY 1141  |||||HASQVCSETPDDLLD|DGETKEDT|SFANDIKESSAVFSKSVQK|GELSRS|PSP|TH|HLAQ 1200
Db 1141  |||||HASQVCSETPDDLLD|DGETKEDT|SFANDIKESSAVFSKSVQK|GELSRS|PSP|TH|HLAQ 1200
QY 1201  |||||GYRRGAKK|LESS|EN|SS|DEEL|PC|FQ|HLLFGKVNNI|PSQSTRHSTVAT|ECLSKNTE|ENL 1260
Db 1201  |||||GYRRGAKK|LESS|EN|SS|DEEL|PC|FQ|HLLFGKVNNI|PSQSTRHSTVAT|ECLSKNTE|ENL 1260
QY 1261  |||||LSLKN|SLNDCSNQVILAKASQEH|LSEETKCSASLFS|SQCSELED|LTANT|QDPFLIGS 1320
Db 1261  |||||LSLKN|SLNDCSNQVILAKASQEH|LSEETKCSASLFS|SQCSELED|LTANT|QDPFLIGS 1320
QY 1321  |||||SKQMRHQS|SQGVGLSDKELVSDDERGTG|LENNQ|EESQMS|N|LGEAASG|CESETSVSE 1380
Db 1321  |||||SKQMRHQS|SQGVGLSDKELVSDDERGTG|LENNQ|EESQMS|N|LGEAASG|CESETSVSE 1380
QY 1381  |||||DCSGLSSQSD|ILTQO|DRTMOHNLK|LQ|E|MA|E|L|V|L|Q|H|G|S|Q|P|S|N|S|Y|P|S|I|S|D|S|A|L|E 1440
Db 1381  |||||DCSGLSSQSD|ILTQO|DRTMOHNLK|LQ|E|MA|E|L|V|L|Q|H|G|S|Q|P|S|N|S|Y|P|S|I|S|D|S|A|L|E 1440
QY 1441  |||||DLRN|PQSTSEKAVLTSQKSSSEYPI|SQNPEGLSADKFEV|SAD|S|T|S|K|N|K|E|P|G|V|E|R|S|S|P|K 1500
Db 1441  |||||DLRN|PQSTSEKAVLTSQKSSSEYPI|SQNPEGLSADKFEV|SAD|S|T|S|K|N|K|E|P|G|V|E|R|S|S|P|K 1500
QY 1501  |||||CPSL|DRWY|HSCSGSLQNRNYP|SQBELIKVVDVEEQ|LEESG|PHDLT|T|S|Y|L|P|ROD|LEG 1560
Db 1501  |||||CPSL|DRWY|HSCSGSLQNRNYP|SQBELIKVVDVEEQ|LEESG|PHDLT|T|S|Y|L|P|ROD|LEG 1560
QY 1561  |||||TPYLESGISL|FSDPDS|EDRAPESARVGN|IP|S|T|S|A|L|K|V|P|Q|L|K|V|A|E|S|Q|P|A|A|H|T 1620
Db 1561  |||||TPYLESGISL|FSDPDS|EDRAPESARVGN|IP|S|T|S|A|L|K|V|P|Q|L|K|V|A|E|S|Q|P|A|A|H|T 1620
QY 1621  |||||DTAGY|NAMEESVSR|EKPELT|ASTERYNKRMSMVV|SGLTPEEFMLVYK|FARKH|HIT|NL|I 1680
Db 1621  |||||DTAGY|NAMEESVSR|EKPELT|ASTERYNKRMSMVV|SGLTPEEFMLVYK|FARKH|HIT|NL|I 1680
QY 1681  |||||TEETHVVMKTD|AEFVCERT|LKYFLGIAGK|VWVSYFWVTQTS|KERKMLNEH|D|F|EVRGDV 1740
Db 1681  |||||TEETHVVMKTD|AEFVCERT|LKYFLGIAGK|VWVSYFWVTQTS|KERKMLNEH|D|F|EVRGDV 1740
QY 1741  |||||VNGRN|HOGK|P|K|A|R|E|S|O|D|R|K|I|F|R|G|L|E|T|C|C|Y|G|P|T|N|M|P|T|D|Q|L|E|W|V|Q|L|C|G|A|S|V|Y|K|E|L|S|S|F|T|L 1800
Db 1741  |||||VNGRN|HOGK|P|K|A|R|E|S|O|D|R|K|I|F|R|G|L|E|T|C|C|Y|G|P|T|N|M|P|T|D|Q|L|E|W|V|Q|L|C|G|A|S|V|Y|K|E|L|S|S|F|T|L 1800
QY 1801  |||||GTGV|HP|V|V|V|V|Q|P|D|A|W|E|D|N|G|F|H|A|I|G|Q|C|E|A|P|V|V|T|R|E|W|L|D|S|V|A|L|Y|Q|C|Q|E|L|D|Y|L|I|P|O|I|P|H 1860
Db 1801  |||||GTGV|HP|V|V|V|V|Q|P|D|A|W|E|D|N|G|F|H|A|I|G|Q|C|E|A|P|V|V|T|R|E|W|L|D|S|V|A|L|Y|Q|C|Q|E|L|D|Y|L|I|P|O|I|P|H 1860
QY 1861  |||||SHY 1863
Db 1861  |||||SHY 1863
```

```

RESULT 14
AAR81529
ID AAR81529 standard; Protein; 1863 AA.
XX
AC AAR81529;
XX
DT 02-OCT-1996 (first entry)
XX
DE BRCA1 mutant from sample set MSK9646.
```

```

XX
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1628 /note= "M1628V"
XX
PN WO9605306-A2.
XX
PD 22-FEB-1996.
XX
PF 11-AUG-1995; 95WO-US10202.
XX
PR 07-JUN-1995; 95US-0483553.
PR 12-AUG-1994; 94US-0289221.
PR 02-SEP-1994; 94US-0300266.
PR 16-SEP-1994; 94US-0308104.
PR 29-NOV-1994; 94US-0348824.
PR 24-MAR-1995; 95US-0409305.
PR 07-JUN-1995; 95US-0480784.
XX
PA (MYRI-) MYRIAD GENETICS INC.
PA (CANC-) CANCER INST.
PA (RECH-) CENT RECH DU CHUL.
XX
PI Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
XX
DR WPI: 1996-139702/14.
DR N-PSDB; AAT17486.
XX
PT New nucleic acid and polypeptide for mutant or polymorphic BRCA1
PT gene - for diagnosis and therapy of human breast and ovarian cancer
PT and for diagnosing pre-disposition to these cancers
XX
PS Claim 1; ; 218pp; English.
XX
CC AAR81483-R81497 and AAR81499-R81546 represent mutations of the protein
CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1)
CC (see AAR81481 for wild type protein). These mutations can be used as
CC immunogens for antibody production. The mutant BRCA1 genes encoding
CC these sequences have at least 1 mutation or polymorphism in comparison
CC to the wild type cDNA (see AAT17438 for wild type). By detecting a
CC germline alteration in the wild type BRCA1 gene, a predisposition for
CC breast and ovarian cancer can be diagnosed. In one method, BRCA1 mRNA
CC isolated from a tissue sample from a subject has a probe, corresponding
CC to a fragment of the cDNA encoding the wild type BRCA1 sequence (or an
CC allele-specific probe for a mutation of it), added to it. The conditions
CC allow for hybridisation of the probe to the mRNA, and any hybridisation
CC which occurs is detected. Alternatively the BRCA1 gene in the tissue
CC sample is isolated, and a shift in electrophoretic mobility of single
CC stranded DNA from the sample on a non-denaturing polyacrylamide gel
CC indicates a mutation. These methods of detection can also diagnose a
CC lesion neoplasia associated with the BRCA1 locus. The methods may be
CC used in gene therapy, protein replacement therapy and protein mimetics,
CC and may be used to screen for drugs in cancer therapy.
XX
SQ Sequence 1863 AA;
Query Match 99.9%; Score 9638; DB 17; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1861; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MDLSALRVEEVQNVINAMQKILECPICLELIKPEVSTKCDHIFCKFCMLKLLNQKKGPSQ 60
Db 1 MDLSALRVEEVQNVINAMQKILECPICLELIKPEVSTKCDHIFCKFCMLKLLNQKKGPSQ 60
QY 61 CPLCKNDITKRSIQESTRFSQVLELLKTKICAFQD|TGL|EYANSY|NFAKKENNSPEHLKD 120
Db 61 CPLCKNDITKRSIQESTRFSQVLELLKTKICAFQD|TGL|EYANSY|NFAKKENNSPEHLKD 120
```


QY 121 EYSIIQSMGYRNRKRLQSEPNPNSLOETSLSVOLSNLGTVRTLRTKORIQPOKTSYVI 180
DB 121 EYSIIQSMGYRNRKRLQSEPNPNSLOETSLSVOLSNLGTVRTLRTKORIQPOKTSYVI 180
QY 181 ELGSSDSEDVANKATYCSVGDQELQITPOGTRDEISLDSAKAAECFSETDVTNTEHHQ 240
DB 181 ELGSSDSEDVANKATYCSVGDQELQITPOGTRDEISLDSAKAAECFSETDVTNTEHHQ 240
QY 241 PSNNDLNTTEKRAAERHPKPYOGSSVSNLHVPCGTNTNTHASSLQHENSSLLLTDRMNYE 300
DB 241 PSNNDLNTTEKRAAERHPKPYOGSSVSNLHVPCGTNTNTHASSLQHENSSLLLTDRMNYE 300
QY 301 KAEFCNKSQKQGLARSHQNRWAGSKETCNDRTPTSTKVKVLDNADPLCKERENKQKLP 360
DB 301 KAEFCNKSQKQGLARSHQNRWAGSKETCNDRTPTSTKVKVLDNADPLCKERENKQKLP 360
QY 361 SENPRDTEVPWITUNSSITQKYNWFSDDELGGSDSDHGESSENAKVAADVLDVNEVD 420
DB 361 SENPRDTEVPWITUNSSITQKYNWFSDDELGGSDSDHGESSENAKVAADVLDVNEVD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVSKSVESNIEDKIFGKTYRKKASLPNLSHYTEN 480
DB 421 EYSGSSEKIDLLASDPHEALICKSERVSKSVESNIEDKIFGKTYRKKASLPNLSHYTEN 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPMINOGTNOTE 540
DB 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPMINOGTNOTE 540
QY 541 QNGQVNNITNSGHEKNTKGDSTQNEKNPNPIESLEKESAFKTKAEPITSSSTSNMELELNI 600
DB 541 QNGQVNNITNSGHEKNTKGDSTQNEKNPNPIESLEKESAFKTKAEPITSSSTSNMELELNI 600
QY 601 HNSKAPKNLRKRSSTRIHIALELVVSRNLSPNCTELOIDSCSSSEIEIKKKYNOMP 660
DB 601 HNSKAPKNLRKRSSTRIHIALELVVSRNLSPNCTELOIDSCSSSEIEIKKKYNOMP 660
QY 661 RHSRLQLMEGREPATGAKSKNPEQTSKRHSDTFPPELKLITNAPGSFTKCSNTSELKE 720
DB 661 RHSRLQLMEGREPATGAKSKNPEQTSKRHSDTFPPELKLITNAPGSFTKCSNTSELKE 720
QY 721 FYNPSLPREEKEKLETYKVSNNADDPKDLMSGERVLTQERSVESSISLVPGTDYGTQ 780
DB 721 FYNPSLPREEKEKLETYKVSNNADDPKDLMSGERVLTQERSVESSISLVPGTDYGTQ 780
QY 781 ESISLLEVSTLGKATEPNKCVSOCAAFENPKGLTHGCSKDNRNDTEGKYPLGHEVNH 840
DB 781 ESISLLEVSTLGKATEPNKCVSOCAAFENPKGLTHGCSKDNRNDTEGKYPLGHEVNH 840
QY 841 RETSIEMEESLDAQYLQNTFKVSKRSQFALFSPNGNAEEECATFSAHSGSLKKQSPKVT 900
DB 841 RETSIEMEESLDAQYLQNTFKVSKRSQFALFSPNGNAEEECATFSAHSGSLKKQSPKVT 900
QY 901 FCEQKEENQGNESNIKPQVTNITAGFPVVGOKDPVDNAKCSIKGGSFRLCSSQFRG 960
DB 901 FCEQKEENQGNESNIKPQVTNITAGFPVVGOKDPVDNAKCSIKGGSFRLCSSQFRG 960
QY 961 NETGLITPNKHGGLQNPYRIPPLPIKSFVTKCKKNLLEENFEHSMSPEREMGNENIP 1020
DB 961 NETGLITPNKHGGLQNPYRIPPLPIKSFVTKCKKNLLEENFEHSMSPEREMGNENIP 1020
QY 1021 STVSTISRNNIRENVFKASSNINEVSGSSITNEIGSSDENTQABLGRNRPGL 1080
DB 1021 STVSTISRNNIRENVFKASSNINEVSGSSITNEIGSSDENTQABLGRNRPGL 1080
QY 1081 NAMRLGLVQPEVYKQSLPGSNCKHPETKQYEVEWQTVNTDFSPYLISDNLBQPMGSS 1140
DB 1081 NAMRLGLVQPEVYKQSLPGSNCKHPETKQYEVEWQTVNTDFSPYLISDNLBQPMGSS 1140
QY 1141 HASQVCSETPDLDLDDGEIKEDTFAENDIKESSAVFSKSVQKGLSRSPSPFTHTHAQ 1200
DB 1141 HASQVCSETPDLDLDDGEIKEDTFAENDIKESSAVFSKSVQKGLSRSPSPFTHTHAQ 1200

QY 1201 GYRRGAKKLESSEENLSEDEELPCFQHLFGKVNNIQSQTRHSTVATECLSKNTEENL 1260
DB 1201 GYRRGAKKLESSEENLSEDEELPCFQHLFGKVNNIQSQTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLUNCDSNQVILAKASQEHHLSEETKCSASFSSQCSELEDLTANTNTQDPFLIGS 1320
DB 1261 LSLKNSLUNCDSNQVILAKASQEHHLSEETKCSASFSSQCSELEDLTANTNTQDPFLIGS 1320
QY 1321 SKQMRHOSQESOGVGLSDKELYSDDDEERCTGLEENNOEQSMDSNLGEAASCESETSVE 1380
DB 1321 SKQMRHOSQESOGVGLSDKELYSDDDEERCTGLEENNOEQSMDSNLGEAASCESETSVE 1380
QY 1381 DCSGLSSQSDILTTQQORDTMQHNLIKLOEMAELAVLEQHGQSQPSNSYPSIISDSSALE 1440
DB 1381 DCSGLSSQSDILTTQQORDTMQHNLIKLOEMAELAVLEQHGQSQPSNSYPSIISDSSALE 1440
QY 1441 DLRNPEGSTSKAVLTQSOKSEYPISQNPEGLSADKFEVSADSDTSKNKEPGEVERSSPSK 1500
DB 1441 DLRNPEGSTSKAVLTQSOKSEYPISQNPEGLSADKFEVSADSDTSKNKEPGEVERSSPSK 1500
QY 1501 CPSLDDRWYMHSCSGSLQNRNYPQOEELIKVVDVEEQOLEESGPHDLTETSYLPQDDLEG 1560
DB 1501 CPSLDDRWYMHSCSGSLQNRNYPQOEELIKVVDVEEQOLEESGPHDLTETSYLPQDDLEG 1560
QY 1561 TPYLESIGISLFSDDPESDRAPE SARVGNIPSSSTSALKVPQLKVAESAQSPAAAHPT 1620
DB 1561 TPYLESIGISLFSDDPESDRAPE SARVGNIPSSSTSALKVPQLKVAESAQSPAAAHPT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEEFMLVYKFAARKHHTLTNLI 1680
DB 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEEFMLVYKFAARKHHTLTNLI 1680
QY 1681 TEETHVVMKTDAEFVCERTLYFLGIAGGKVVYSYFNWVTOSIKERKMLNEHDFEVRGDV 1740
DB 1681 TEETHVVMKTDAEFVCERTLYFLGIAGGKVVYSYFNWVTOSIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHGPKRARESQRKIFRGLICCYGPTNNPTDQLEMMVOLCGASVVKELSSFTL 1800
DB 1741 VNGRNHGPKRARESQRKIFRGLICCYGPTNNPTDQLEMMVOLCGASVVKELSSFTL 1800
QY 1801 GTGVHPVIVVQPDWATEDNGPHAIQOMCEAPVVTREWVLDSDVALYQCQELDTYLIPIPH 1860
DB 1801 GTGVHPVIVVQPDWATEDNGPHAIQOMCEAPVVTREWVLDSDVALYQCQELDTYLIPIPH 1860
QY 1861 SHY 1863
DB 1861 SHY 1863
RESULT 15
AAR81532
ID AAR81532 standard; Protein; 1863 AA.
XX AAR81532;
AC XX
DT 02-OCT-1996 (first entry)
TX
DE BRCA1 mutant from sample set MSK7542.
XX
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;
KW antibody production; germline alteration; probe; lesion neoplasia; human;
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1852
FT /note= "r1852s"
XX
PN W09605306-A2.
XX
PD 22-FEB-1996.
XX

PF 11-AUG-1995; 95W0-US10202.
XX
PR 07-JUN-1995; 95US-0483553.
PR 12-AUG-1994; 94US-0289221.
PR 02-SEP-1994; 94US-0300266.
PR 16-SEP-1994; 94US-0308104.
PR 29-NOV-1994; 94US-0348824.
PR 24-MAR-1995; 95US-0409305.
PR 07-JUN-1995; 95US-0480784.
XX
PA (MYRI-) MYRIAD GENETICS INC.
PA (CANC-) CANCER INST.
PA (RECH-) CENT RECH DU CHUL.
XX
PI Durocher F, Emi M, Nakamura Y, Simard J, Shattuck-Eidens DM;
XX
DR WPI: 1996-139702/14.
DR N-PSDB; AAT17489.
XX
PT New nucleic acid and polypeptide for mutant or polymorphic BRCA1
PT gene - for diagnosis and therapy of human breast and ovarian cancer
PT and for diagnosing pre-disposition to these cancers
PT
XX
PS Claim 1; : 218pp; English.
XX
CC AAR81483-R81497 and AAR81499-R81546 represent mutations of the protein
CC encoded by the human breast and ovarian cancer predisposing gene (BRCA1
CC (see AAR81481 for wild type protein). These mutations can be used as
CC immunogens for antibody production. The mutant BRCA1 genes encoding
CC these sequences have at least 1 mutation or polymorphism in comparison
CC to the wild type cDNA (see AAT17438 for wild type). By detecting a
CC germline alteration in the wild type BRCA1 gene, a predisposition for
CC breast and ovarian cancer can be diagnosed. In one method, BRCA1 mRNA
CC isolated from a tissue sample from a subject has a probe, corresponding
CC to a fragment of the cDNA encoding the wild type BRCA1 sequence (or an
CC allele-specific probe for a mutation of it), added to it. The conditions
CC allow for hybridisation of the probe to the mRNA, and any hybridisation
CC which occurs is detected. Alternatively the BRCA1 gene in the tissue
CC sample is isolated, and a shift in electrophoretic mobility of single
CC stranded DNA from the sample on a non-denaturing polyacrylamide gel
CC indicates a mutation. These methods of detection can also diagnose a
CC lesion neoplasia associated with the BRCA1 locus. The methods may be
CC used in gene therapy, protein replacement therapy and protein mimetics,
CC and may be used to screen for drugs in cancer therapy.
XX
SQ Sequence 1863 AA;

Query Match 99.9%; Score 9638; DB 17; Length 1863;
Best Local Similarity 99.9%; Pred. NO. 0;
Matches 1861; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDLSALRVEEVQVINAMQKILECPICLELIKPEVSTKCDHIFCKFCMLKLLNOKKGPQS 60
DB 1 MDLSALRVEEVQVINAMQKILECPICLELIKPEVSTKCDHIFCKFCMLKLLNOKKGPQS 60

QY 61 CPLCKNDITKRSIQESTRFSQVVEELKIIICAFQDGTGLGYANSYNPAKKENNSPEHLKD 120
DB 61 CPLCKNDITKRSIQESTRFSQVVEELKIIICAFQDGTGLGYANSYNPAKKENNSPEHLKD 120

QY 121 EYSLIQSMGYRNRAKRLQSEPNPSLQETSLSVQLSNLGTVRTLRKQIQPKTSVYI 180
DB 121 EYSLIQSMGYRNRAKRLQSEPNPSLQETSLSVQLSNLGTVRTLRKQIQPKTSVYI 180

QY 181 ELGSDSESDTVNKKATCSVGDQELLOITPGOTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
DB 181 ELGSDSESDTVNKKATCSVGDQELLOITPGOTRDEISLDSAKKAACEFSETDVTNTEHHQ 240

QY 241 PSNNDLNTEKRAAERHPEKYQSSVSNLHVPCGNTNTHASSLQHENSILLTKDRMNVE 300
DB 241 PSNNDLNTEKRAAERHPEKYQSSVSNLHVPCGNTNTHASSLQHENSILLTKDRMNVE 300

QY 301 KAEFCNKSQOGLARSOHNWAGSKETCNDRRTPSTPEKKVVDLADPLCERKEWNKQKLPC 360
DB 301 KAEFCNKSQOGLARSOHNWAGSKETCNDRRTPSTPEKKVVDLADPLCERKEWNKQKLPC 360

DB 301 KAEFCNKSQOGLARSOHNWAGSKETCNDRRTPSTPEKKVVDLADPLCERKEWNKQKLPC 360
QY 361 SENPRDTEVPWITTLNSSIOKVNEWFSRSDLLGSDSDHDSGESNAKAVADVLDVLENYD 420
DB 361 SENPRDTEVPWITTLNSSIOKVNEWFSRSDLLGSDSDHDSGESNAKAVADVLDVLENYD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVBESNIEDKIFGKTYRKASLPNLSHVTEN 480
DB 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVBESNIEDKIFGKTYRKASLPNLSHVTEN 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPMINOGNQT 540
DB 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPMINOGNQT 540
QY 541 QNGOVNMITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTAEPISSSISNWELELNI 600
DB 541 QNGOVNMITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTAEPISSSISNWELELNI 600
QY 601 HNSKAPKKNLRKRKSTRHIHALELVVSRNLSPNCTELQIDSCSSSEETKKKYNOMPV 660
DB 601 HNSKAPKKNLRKRKSTRHIHALELVVSRNLSPNCTELQIDSCSSSEETKKKYNOMPV 660
QY 661 RHSRNLQLMGEGKEPATGAKKSNKPNQOTSKRHSDDTFPELKLTPNAGSFTKCSNTSELKE 720
DB 661 RHSRNLQLMGEGKEPATGAKKSNKPNQOTSKRHSDDTFPELKLTPNAGSFTKCSNTSELKE 720
QY 721 FVNPSPREEKEKLETVKYSNNAEDPKDLMLSGERVLOTERSVESSSISLVPGTDYGTQ 780
DB 721 FVNPSPREEKEKLETVKYSNNAEDPKDLMLSGERVLOTERSVESSSISLVPGTDYGTQ 780
QY 781 ESISLLEVSTLGAKTEPNKCVSQCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHEVNH 840
DB 781 ESISLLEVSTLGAKTEPNKCVSQCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHEVNH 840
QY 841 RETSIEMESELDAQYLQNTFKVSKROSFPALFSPNPGNAEBECATFSAHSGSLKQSPKV 900
DB 841 RETSIEMESELDAQYLQNTFKVSKROSFPALFSPNPGNAEBECATFSAHSGSLKQSPKV 900
QY 901 FECEQKEENGKGNESNPKPVOTVNTAGFPVVGOKDPVDNAKCSIKGSRFCLSSQFRG 960
DB 901 FECEQKEENGKGNESNPKPVOTVNTAGFPVVGOKDPVDNAKCSIKGSRFCLSSQFRG 960
QY 961 NETGLITPNKHGLLQNPYRIPPLPKSFVKTCKKNLLEENFEHSMSPEREMGNIP 1020
DB 961 NETGLITPNKHGLLQNPYRIPPLPKSFVKTCKKNLLEENFEHSMSPEREMGNIP 1020
QY 1021 STVSTISRNNIRENVFKEASSSNI NEVGSTNEVGSSINEIGSSDENIQAEIAGNRGPKL 1080
DB 1021 STVSTISRNNIRENVFKEASSSNI NEVGSTNEVGSSINEIGSSDENIQAEIAGNRGPKL 1080
QY 1081 NAMLRGLVLOPEVYKOSLPGSNCKHPEIKKOEYEEVVQTVNTDFSPYLIISDNLEQPMGSS 1140
DB 1081 NAMLRGLVLOPEVYKOSLPGSNCKHPEIKKOEYEEVVQTVNTDFSPYLIISDNLEQPMGSS 1140
QY 1141 HASQVCSETPDDLDDGEIKEDTSFAENDIKESSAVFSKVQKGLSRSPSPFTHHLAQ 1200
DB 1141 HASQVCSETPDDLDDGEIKEDTSFAENDIKESSAVFSKVQKGLSRSPSPFTHHLAQ 1200
QY 1201 GYRRGAKKLESSSENLSSEDEELPCFQHLFLGKVNINIPQSOTRHSVTATECLSKNTEENL 1260
DB 1201 GYRRGAKKLESSSENLSSEDEELPCFQHLFLGKVNINIPQSOTRHSVTATECLSKNTEENL 1260
QY 1261 LSLKNSLNDCSNOVILAKASQEHHLSEETKCSASLFSQCSSELEDTANTNTODPFLIGS 1320
DB 1261 LSLKNSLNDCSNOVILAKASQEHHLSEETKCSASLFSQCSSELEDTANTNTODPFLIGS 1320
QY 1321 SKQMRHQSESQGVGLSKDELVSDEERGTLGLENNQEQMSDNLGEAASGCSESETSVSE 1380
DB 1321 SKQMRHQSESQGVGLSKDELVSDEERGTLGLENNQEQMSDNLGEAASGCSESETSVSE 1380
QY 1381 DCSGLSSQSDILITQOQDRTMQHNLIKLQEQMAEILEAVLEQHGQSPNSYPSIISDSSALE 1440
DB 1381 DCSGLSSQSDILITQOQDRTMQHNLIKLQEQMAEILEAVLEQHGQSPNSYPSIISDSSALE 1440

Qy	1441	DLRNEQSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSSTSKNKEPGVERSSPSK	1500
Db	1441	DLRNEQSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSSTSKNKEPGVERSSPSK	1500
Qy	1501	CPSLDDRMYHSCSGSLQNRNYPQBELIKVDVEEQLEESGPHDLTETSYLPRODLEG	1560
Db	1501	CPSLDDRMYHSCSGSLQNRNYPQBELIKVDVEEQLEESGPHDLTETSYLPRODLEG	1560
Qy	1561	TPYLESGISLFSDDPESDPESEDRAPE SARVGNIPSSTSALKVPQLKVAESAQSPAAHHT	1620
Db	1561	TPYLESGISLFSDDPESDPESEDRAPE SARVGNIPSSTSALKVPQLKVAESAQSPAAHHT	1620
Qy	1621	DTAGYNAMEESVSREKPELTASTERYNKRMVMVSGLTPEEFMLVYKFARKHHITLNL	1680
Db	1621	DTAGYNAMEESVSREKPELTASTERYNKRMVMVSGLTPEEFMLVYKFARKHHITLNL	1680
Qy	1681	TEETHVVMKTDAEFVCERTLKYFLGIAGGKVVVSYFWVTQSIKERKMLNEHDFEVRGDV	1740
Db	1681	TEETHVVMKTDAEFVCERTLKYFLGIAGGKVVVSYFWVTQSIKERKMLNEHDFEVRGDV	1740
Qy	1741	VNGRNHGGPKRARESDRKIFRGLEICCYGPTNNMPTDQLEWVQLCGASVVKELSSFTL	1800
Db	1741	VNGRNHGGPKRARESDRKIFRGLEICCYGPTNNMPTDQLEWVQLCGASVVKELSSFTL	1800
Qy	1801	GTGVHPITVVVQPDWNTEDNGFHAIGOMCEAPVVTREWLDSVALYQCQELDTYLIPQIPH	1860
Db	1801	GTGVHPITVVVQPDWNTEDNGFHAIGOMCEAPVVTREWLDSVALYQCQELDTYLIPQIPH	1860
Qy	1861	SHY 1863	
Db	1861	SHY 1863	

Search completed: January 22, 2003, 16:46:04
Job time : 76 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 16:46:04 ; Search time 15.5 Seconds
(without alignments)
2425.332 Million cell updates/sec

Title: US-09-734-672-4

Perfect score: 9649

Sequence: 1 MDLSALRVEEVQNVINAMQK.....LYQCQLDYLIPQIPHSY 1863

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 122226 seqs, 2017851 residues

Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	9649	100.0	1863	9 US-09-734-672-4	Sequence 4, Appli
2	9635	99.9	1863	9 US-09-734-672-2	Sequence 2, Appli
3	9635	99.9	1863	9 US-09-734-672-6	Sequence 6, Appli
4	357	3.7	2344	10 US-09-815-242-12713	Sequence 12713, A
5	341	3.5	6281	10 US-09-815-242-12996	Sequence 12996, A
6	328.5	3.4	2368	10 US-09-815-242-5635	Sequence 5635, Ap
7	328.5	3.4	2368	10 US-09-815-242-12389	Sequence 12389, A
8	314	3.3	1596	9 US-09-902-432-4	Sequence 4, Appli
9	312	3.2	2478	10 US-09-764-176-7	Sequence 7, Appli
10	312	3.2	2478	10 US-09-815-242-5816	Sequence 5816, Ap
11	312	3.2	2478	10 US-09-815-242-12967	Sequence 12967, A
12	308	3.2	2665	10 US-09-864-761-34248	Sequence 34248, A
13	306	3.2	3158	10 US-09-815-242-12611	Sequence 12611, A
14	299.5	3.1	2843	9 US-09-987-482-1	Sequence 1, Appli
15	298.5	3.1	2843	8 US-08-681-219-32	Sequence 32, Appli
16	286.5	3.0	1781	10 US-09-738-877-3	Sequence 3, Appli
17	285.5	3.0	2025	10 US-09-815-242-5703	Sequence 5703, Ap
18	282	2.9	50	10 US-09-998-667-15	Sequence 15, Appl
19	278	2.9	3256	10 US-09-919-172-98	Sequence 98, Appl

20	272	2.8	2437	10 US-09-815-242-5834	Sequence 5834, Ap
21	268	2.8	5795	10 US-09-815-242-12610	Sequence 12610, A
22	261	2.7	1285	10 US-09-982-091A-2	Sequence 2, Appli
23	258	2.7	1786	9 US-09-742-096-3	Sequence 3, Appli
24	255	2.6	2434	10 US-09-815-242-5835	Sequence 5835, Ap
25	253.5	2.6	1269	10 US-09-815-242-13113	Sequence 13113, A
26	251.5	2.6	1325	10 US-09-864-761-35612	Sequence 35612, A
27	251	2.6	2125	10 US-09-919-172-29	Sequence 29, Appl
28	250.5	2.6	1501	10 US-09-924-154-17	Sequence 17, Appl
29	250.5	2.6	2086	10 US-09-815-242-5639	Sequence 5639, Ap
30	250	2.6	2789	10 US-09-801-574-57	Sequence 57, Appl
31	248.5	2.6	2139	10 US-09-727-384-6	Sequence 6, Appli
32	248	2.6	1215	10 US-09-815-242-5908	Sequence 5908, Ap
33	241.5	2.5	1111	10 US-09-815-242-12955	Sequence 12955, A
34	235.5	2.4	1332	10 US-09-982-091A-4	Sequence 4, Appli
35	235	2.4	48	10 US-09-864-761-37757	Sequence 37757, A
36	234.5	2.4	1618	9 US-09-963-875-1	Sequence 1, Appli
37	234.5	2.4	1884	10 US-09-785-770A-17	Sequence 17, Appl
38	234.5	2.4	1907	10 US-09-785-770A-16	Sequence 16, Appl
39	232.5	2.4	2548	10 US-09-851-682A-1	Sequence 1, Appli
40	231.5	2.4	1346	9 US-09-902-432-2	Sequence 2, Appli
41	230.5	2.4	2308	12 US-10-000-954-2	Sequence 2, Appli
42	227.5	2.4	1086	10 US-09-924-154-15	Sequence 15, Appl
43	227.5	2.4	1597	9 US-09-832-292-35	Sequence 35, Appl
44	226.5	2.3	2785	10 US-09-801-574-8	Sequence 8, Appli
45	219.5	2.3	1478	10 US-09-801-368-52	Sequence 52, Appl

ALIGNMENTS

RESULT 1

US-09-734-672-4

; Sequence 4, Application US/09734672

; Publication No. US20020183268A1

; GENERAL INFORMATION:

; APPLICANT: Murphy, Patricia D.

; Allen, Antoinette C.

; Alvarres, Christopher P.

; Critz, Brenda S.

; Olson, Sheri J.

; Schelter, Denise B.

; Zeng, Bin

; TITLE OF INVENTION: Coding Sequences of the Human

; BRCA1 Gene

; NUMBER OF SEQUENCES: 72

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Morgan Lewis & Bockius LLP

; STREET: 1111 Pennsylvania Ave., N.W.

; CITY: Washington

; STATE: District of Columbia

; COUNTRY: USA

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA: US/09/734,672

; FILING DATE: 03-Dec-2000

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 08/966,436

; FILING DATE: 07-No. US20020183268A1-97

; APPLICATION NUMBER: US 08/598,591

; FILING DATE: 12-Feb-96

; ATTORNEY/AGENT INFORMATION:

; NAME: Michael S. Tuscan

; REGISTRATION NUMBER: 43,210

; REFERENCE/DOCKET NUMBER: 44921-5055-02-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-739-3000

; TELEFAX: 202-739-3001

; INFORMATION FOR SEQ ID NO: 4:									
; SEQUENCE CHARACTERISTICS:									
; LENGTH: 1863 amino acids									
; TYPE: amino acid									
; STRANDEDNESS: No. US20020183268A1 Relevant									
; TOPOLOGY: No. US20020183268A1 Relevant									
; MOLECULE TYPE: protein									
; ORIGINAL SOURCE:									
; ORGANISM: Homo sapiens									
; STRAIN: BRCA1									
; POSITION IN GENOME:									
; CHROMOSOME/SEGMENT: 17									
; MAP POSITION: 17q21									
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:									
US-09-734-672-4									
Query Match 100.0%; Score 9649; DB 9; Length 1863;									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 1863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MDLSALRVEEVQVNVINAMOKITLPCICILELIKPEVSTKCDHIFCKFCMLKLLNOKKGPSQ	60						
DB	1	MDLSALRVEEVQVNVINAMOKITLPCICILELIKPEVSTKCDHIFCKFCMLKLLNOKKGPSQ	60						
QY	61	CPLCKNDITKRSLOSTRFSQLVEELKIIICAFQDITGLEYANSYNFAKKENNSPEHLKD	120						
DB	61	CPLCKNDITKRSLOSTRFSQLVEELKIIICAFQDITGLEYANSYNFAKKENNSPEHLKD	120						
QY	121	EVSIIITQSGYRNRARLLQSPENPSLOETSLVSOLSNLGTVRTLRKTRQPOKTSYVI	180						
DB	121	EVSIIITQSGYRNRARLLQSPENPSLOETSLVSOLSNLGTVRTLRKTRQPOKTSYVI	180						
QY	181	ELGSDSEDTVNKATYCSVGQDQLLOITPOGTRDEISLDSAKKAACEFSEDTVNTTEHHQ	240						
DB	181	ELGSDSEDTVNKATYCSVGQDQLLOITPOGTRDEISLDSAKKAACEFSEDTVNTTEHHQ	240						
QY	241	PSNNDLNTERRAAERHPEKYGQSSVSNLHVPCGTNTHASSLQHENSILLTKDRMNV	300						
DB	241	PSNNDLNTERRAAERHPEKYGQSSVSNLHVPCGTNTHASSLQHENSILLTKDRMNV	300						
QY	301	KAFCNKSQKQGLARSHNRWAGSKETCNDRRTPSTKTKVDLNDADPLCERKENWKQLPC	360						
DB	301	KAFCNKSQKQGLARSHNRWAGSKETCNDRRTPSTKTKVDLNDADPLCERKENWKQLPC	360						
QY	361	SENPRDTEVPWITLNSIIQKVNWFERSDELGSDSDSHDGESESNKAVADVLVNEVD	420						
DB	361	SENPRDTEVPWITLNSIIQKVNWFERSDELGSDSDSHDGESESNKAVADVLVNEVD	420						
QY	421	EYSGSSEKIDLLADPHALICKSERVHKSVESNIEDKIFGKTYRKKASLPNLSHVTEN	480						
DB	421	EYSGSSEKIDLLADPHALICKSERVHKSVESNIEDKIFGKTYRKKASLPNLSHVTEN	480						
QY	481	LIIGAFVTEPQIIQERPLTNLKKRRRPTSGLHPEDFIKKADLAVQKTPEMINGTNOTE	540						
DB	481	LIIGAFVTEPQIIQERPLTNLKKRRRPTSGLHPEDFIKKADLAVQKTPEMINGTNOTE	540						
QY	541	ONGQVNNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNMELELNI	600						
DB	541	ONGQVNNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNMELELNI	600						
QY	601	HNSKAPKNRLRRKSSSTRHIIHAELVVSRLNSPPNCTELQIDSCSSSEELKKKKYNOMP	660						
DB	601	HNSKAPKNRLRRKSSSTRHIIHAELVVSRLNSPPNCTELQIDSCSSSEELKKKKYNOMP	660						
QY	661	RHSRLQIMEGKEPATGAKSKNKNPNEOTSKRHSDDTPELKLTNAPGSFTKCSNTSELKE	720						
DB	661	RHSRLQIMEGKEPATGAKSKNKNPNEOTSKRHSDDTPELKLTNAPGSFTKCSNTSELKE	720						
QY	721	FVNPSLPREEKEELKLVKVSNNAAEDPKDMLSGERVLOQTSESSSISLVPGTDYGTQ	780						
DB	721	FVNPSLPREEKEELKLVKVSNNAAEDPKDMLSGERVLOQTSESSSISLVPGTDYGTQ	780						
QY	781	ESISLLEVSTLGKAKTEPNKCVSOCAAFENPKGLIHGCSKDNRNDETEGFKYPLGHEVNH	840						

DB	781	ESISLLEVSTLGKAKTEPNKCVSOCAAFENPKGLIHGCSKDNRNDETEGFKYPLGHEVNH	840						
QY	841	RETSIEMEESLDAQYLQNTFKVSKRQSFALFSPGNABEBCATFSAHSGSLKKQSPKVT	900						
DB	841	RETSIEMEESLDAQYLQNTFKVSKRQSFALFSPGNABEBCATFSAHSGSLKKQSPKVT	900						
QY	901	FECQKEENOGKNESNPKVOTVNITAGFPVVGOKDKPVDNAKCSIKGGSRFCSSQPRG	960						
DB	901	FECQKEENOGKNESNPKVOTVNITAGFPVVGOKDKPVDNAKCSIKGGSRFCSSQPRG	960						
QY	961	NETGLITPNKHGLLQNPYRIPPLPIKSFVTKCKKNLLEENFEHSHMSPEREMGENIIP	1020						
DB	961	NETGLITPNKHGLLQNPYRIPPLPIKSFVTKCKKNLLEENFEHSHMSPEREMGENIIP	1020						
QY	1021	STVSTISRNNIRENVFKEASSNINEVGSSNINEGSSNINEGSSDENTOAELGRNRGPKL	1080						
DB	1021	STVSTISRNNIRENVFKEASSNINEVGSSNINEGSSNINEGSSDENTOAELGRNRGPKL	1080						
QY	1081	NAMRLGLVLOPEVYKQSLPGSNCKHPETIKQOYBEVVQTVNTDFSPYLISDNLEQPMGSS	1140						
DB	1081	NAMRLGLVLOPEVYKQSLPGSNCKHPETIKQOYBEVVQTVNTDFSPYLISDNLEQPMGSS	1140						
QY	1141	HASQVCSETPDDLLDDGEIKEDTSAFENDIKESSAVFSKQKGLSRSPSPFTHHLAQ	1200						
DB	1141	HASQVCSETPDDLLDDGEIKEDTSAFENDIKESSAVFSKQKGLSRSPSPFTHHLAQ	1200						
QY	1201	GYRGAKKLESSEENLSEDEELPCFQHLHFGKVNINIPQSSTRHSTVATECLSKNTEENL	1260						
DB	1201	GYRGAKKLESSEENLSEDEELPCFQHLHFGKVNINIPQSSTRHSTVATECLSKNTEENL	1260						
QY	1261	LSLKNLSNDCSNQVILAKASQEHLSBETKCSASFSSQCSLELDTANTQDPFLIGS	1320						
DB	1261	LSLKNLSNDCSNQVILAKASQEHLSBETKCSASFSSQCSLELDTANTQDPFLIGS	1320						
QY	1321	SKQMRHOSQSGVGLSDKELVSDDEERGCTGLEENNQBESQSDSNLGEAASCESETSVSE	1380						
DB	1321	SKQMRHOSQSGVGLSDKELVSDDEERGCTGLEENNQBESQSDSNLGEAASCESETSVSE	1380						
QY	1381	DCSGLSSQSDILTTQQRDTMOHNLIKQOEMAELEAVLEHQHQSOPSNYSYPSIISSSAALE	1440						
DB	1381	DCSGLSSQSDILTTQQRDTMOHNLIKQOEMAELEAVLEHQHQSOPSNYSYPSIISSSAALE	1440						
QY	1441	DLRPEOSTSEKAVLTQKSEXPISQNPGLSADKPEVSADSSTSKNKEPCGVERSSPSK	1500						
DB	1441	DLRPEOSTSEKAVLTQKSEXPISQNPGLSADKPEVSADSSTSKNKEPCGVERSSPSK	1500						
QY	1501	CPSLDDRWYHSCSGSLQNRNYPQOEELIKVVDVEEQQLBESGPHDLTETSYLPRQDLEG	1560						
DB	1501	CPSLDDRWYHSCSGSLQNRNYPQOEELIKVVDVEEQQLBESGPHDLTETSYLPRQDLEG	1560						
QY	1561	TPYLESGISLFSDDPESDSDRAPESARVGNIPSSTSALKVPQLKVAESAQSPAAAHTT	1620						
DB	1561	TPYLESGISLFSDDPESDSDRAPESARVGNIPSSTSALKVPQLKVAESAQSPAAAHTT	1620						
QY	1621	DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEEFMLVYKFAKHHTLTNLI	1680						
DB	1621	DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEEFMLVYKFAKHHTLTNLI	1680						
QY	1681	TEETHVVMKTDAEFVCGERTLKYFLGAGGKVVVSYFWVTQSIKERKMLNHDHFEVRGDV	1740						
DB	1681	TEETHVVMKTDAEFVCGERTLKYFLGAGGKVVVSYFWVTQSIKERKMLNHDHFEVRGDV	1740						
QY	1741	VNGRNHGQPKRARESQRKIPRGLIEICCYGPTTNPMTDQLEMMVOLCGASVVKELSSFTL	1800						
DB	1741	VNGRNHGQPKRARESQRKIPRGLIEICCYGPTTNPMTDQLEMMVOLCGASVVKELSSFTL	1800						
QY	1801	GTGVHPITVVQPDAAWTEEDNGFHAIGQMCAPVTVREWVLDVALYQCOELDTYLPQIPH	1860						
DB	1801	GTGVHPITVVQPDAAWTEEDNGFHAIGQMCAPVTVREWVLDVALYQCOELDTYLPQIPH	1860						
QY	1861	SHY 1863							

Db 1861 SHY 1863

RESULT 2

US-09-734-672-2
; Sequence 2, Application US/09734672
; Publication No. US20020183268A1
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; Allen, Antonette C.
; Alvares, Christopher P.
; Critz, Brenda S.
; Olson, Sheri J.
; Schelter, Denise B.
; zeng, Bin
; TITLE OF INVENTION: Coding Sequences of the Human
; BRCAL Gene
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan Lewis & Bockius LLP
; STREET: 1111 Pennsylvania Ave., N.W.
; CITY: Washington
; STATE: District of Columbia
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/734,672
; FILING DATE: 03-Dec-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/966,436
; FILING DATE: 07-No. US20020183268A1-97
; APPLICATION NUMBER: US 08/598,591
; FILING DATE: 12-Feb-96
; ATTORNEY/AGENT INFORMATION:
; NAME: Michael S. Tuscan
; REGISTRATION NUMBER: 43,210
; REFERENCE/DOCKET NUMBER: 44921-5055-02-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-739-3000
; TELEFAX: 202-739-3001
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: No. US20020183268A1 Relevant
; TOPOLOGY: No. US20020183268A1 Relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCAL
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17
; MAP POSITION: 17q21
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-734-672-2

Query Match 99.9%; Score 9635; DB 9; Length 1863;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1860; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MDLSALRVEEVQNVINAMQKILECPICLELIKPEVSTKCDHIKFCMKLLNQKGPSQ 60
Db 1 MDLSALRVEEVQNVINAMQKILECPICLELIKPEVSTKCDHIKFCMKLLNQKGPSQ 60
Qy 61 CPLCKNDITKRSLOESTRFSQVLEELKIIICAFQDITGLGYANSYNFAKKYENNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRFSQVLEELKIIICAFQDITGLGYANSYNFAKKYENNSPEHLKD 120

QY 121 EVSIIQSMGYRNRAKRLQLQSEPNPSLQETSLSVQLSNLGTVRTLRKQRIQPKQTSVYI 180
Db 121 EVSIIQSMGYRNRAKRLQLQSEPNPSLQETSLSVQLSNLGTVRTLRKQRIQPKQTSVYI 180
QY 181 ELGSDSSSDTVNKATYCSVGDQELLQITPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSSDTVNKATYCSVGDQELLQITPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNDLNTTEKRAAERHPEKYQGVSSVSNLHVEPCGNTTHASSLQHENSLLLTQDRMNVE 300
Db 241 PSNDLNTTEKRAAERHPEKYQGVSSVSNLHVEPCGNTTHASSLQHENSLLLTQDRMNVE 300
QY 301 KAEFCNKSKQPLARSQHNRWAGSKETCNDRRTPSTPEKKYDLNADPLCERKEWNKQLPC 360
Db 301 KAEFCNKSKQPLARSQHNRWAGSKETCNDRRTPSTPEKKYDLNADPLCERKEWNKQLPC 360
QY 361 SENPRDTEVPWITLSSSIQKVNEWFSRDELGLGSDSDHGESESNKAVADVLVLENYD 420
Db 361 SENPRDTEVPWITLSSSIQKVNEWFSRDELGLGSDSDHGESESNKAVADVLVLENYD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKKASLPNLSHVTE 480
Db 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKKASLPNLSHVTE 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVOKTPEMINQGTNOTE 540
Db 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVOKTPEMINQGTNOTE 540
QY 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKSAFKTKAEPISSSISNWELELNI 600
Db 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKSAFKTKAEPISSSISNWELELNI 600
QY 601 HNSKAPKKNRLRRKSTRHIALELVVSRNLSPNCTELOIDSCSSSEELTKKKYNOMPV 660
Db 601 HNSKAPKKNRLRRKSTRHIALELVVSRNLSPNCTELOIDSCSSSEELTKKKYNOMPV 660
QY 661 RHSRNLQMEGKEPATGAKKSNKPNQETSKRHDSDTFPELKLTNAPGSFTKCSNTSELKE 720
Db 661 RHSRNLQMEGKEPATGAKKSNKPNQETSKRHDSDTFPELKLTNAPGSFTKCSNTSELKE 720
QY 721 FVNPSPREEKEKLETIVKYSNNAEDPKDMLSGERVLOQTORSVESSISLVPGTDYGTQ 780
Db 721 FVNPSPREEKEKLETIVKYSNNAEDPKDMLSGERVLOQTORSVESSISLVPGTDYGTQ 780
QY 781 ESTSLLEVSTLGRAKTEPNKCVSQAAFPENPKGLIHGCCSDNRNDTEGFKYPLGHEVNH 840
Db 781 ESTSLLEVSTLGRAKTEPNKCVSQAAFPENPKGLIHGCCSDNRNDTEGFKYPLGHEVNH 840
QY 841 RETSIEMESELDAQYLQNTFKVSKRQSFALFSPNGNAEECATFSAHSGSLKKQSPKVT 900
Db 841 RETSIEMESELDAQYLQNTFKVSKRQSFALFSPNGNAEECATFSAHSGSLKKQSPKVT 900
QY 901 FECEQKEENOGKNESNIKPVQTNITAGFPVVGOKDPVDNAKCSIKGSRFCFLSSQFRG 960
Db 901 FECEQKEENOGKNESNIKPVQTNITAGFPVVGOKDPVDNAKCSIKGSRFCFLSSQFRG 960
QY 961 NETGLIIPNKHGILLQNPYRIPPLFPKISFKVTKCKNLEENFEHSHMSPEREMGNINIP 1020
Db 961 NETGLIIPNKHGILLQNPYRIPPLFPKISFKVTKCKNLEENFEHSHMSPEREMGNINIP 1020
QY 1021 STVSTISRNNIRENVFKAEASSNINEVGSSTNEVGSSTNEIGSSDENIQALGNRGPKL 1080
Db 1021 STVSTISRNNIRENVFKAEASSNINEVGSSTNEVGSSTNEIGSSDENIQALGNRGPKL 1080
QY 1081 NAMLRGLVQPEVYKQSLPGSNCKHPKIKQOEYEEVQVTNTDFSPYLIISDNLEQPMGSS 1140
Db 1081 NAMLRGLVQPEVYKQSLPGSNCKHPKIKQOEYEEVQVTNTDFSPYLIISDNLEQPMGSS 1140
QY 1141 HASQVCSETPDLLDDGGEIKEDTSAFENDIKESSAVFSKVQKGELSRSPPTHTHLAQ 1200
Db 1141 HASQVCSETPDLLDDGGEIKEDTSAFENDIKESSAVFSKVQKGELSRSPPTHTHLAQ 1200
QY 1201 GYRRGAKKLESSEENLSSEDEELPCFQHLLFGKVNINIPSQSTRHSTVATECLSKNTEENL 1260

Db 1201 GYRRGAKLSESEENLSSDEELPCFQHLFLCKVNNIPSQSRHSTVATECLSKKTEENL 1260
QY 1261 LSLKSLNDCSNQVILAKASQEHLLSEETKCSASLFSSQCSLEDLTANTNTQDPFLTGS 1320
Db 1261 LSLKSLNDCSNQVILAKASQEHLLSEETKCSASLFSSQCSLEDLTANTNTQDPFLTGS 1320
QY 1321 SKOMRHQSESQGVGLSDKELVDDEERGTLLENNOEQSDNSLGEAASGCESETSYSE 1380
Db 1321 SKOMRHQSESQGVGLSDKELVDDEERGTLLENNOEQSDNSLGEAASGCESETSYSE 1380
QY 1381 DCSGLSSQSDILTQQQRTDMQHNLKLQEMAELEAVLEQHGQSQSPNSYPSIISDSSALE 1440
Db 1381 DCSGLSSQSDILTQQQRTDMQHNLKLQEMAELEAVLEQHGQSQSPNSYPSIISDSSALE 1440
QY 1441 DLNRPQSTSEKAVLTSQKSSEYPISONPEGLSADKFVSADSTSKNKEPVERSSPSK 1500
Db 1441 DLNRPQSTSEKAVLTSQKSSEYPISONPEGLSADKFVSADSTSKNKEPVERSSPSK 1500
QY 1501 CPSLDRRWYMHSCGSLQNRNYPQOEELIKVVDVEEQOLEESGPHDLTETSYLPRQDLEG 1560
Db 1501 CPSLDRRWYMHSCGSLQNRNYPQOEELIKVVDVEEQOLEESGPHDLTETSYLPRQDLEG 1560
QY 1561 TPYLESGISLFDPSDPSSEDRAPEARVGNIPSTTSALKVPQLKVAESAGQPAAAHTT 1620
Db 1561 TPYLESGISLFDPSDPSSEDRAPEARVGNIPSTTSALKVPQLKVAESAGQPAAAHTT 1620
QY 1621 DTAGYNAMESVSRKPELTASTERVNRKMSVMVSGLTPEEFMLVYKFAKHHITLTNLI 1680
Db 1621 DTAGYNAMESVSRKPELTASTERVNRKMSVMVSGLTPEEFMLVYKFAKHHITLTNLI 1680
QY 1681 TTEETHVMKTDAEFCERTLKYFLGIAGGKVVVSYFWWTQSIKERKMLNEHDFEVRGDV 1740
Db 1681 TTEETHVMKTDAEFCERTLKYFLGIAGGKVVVSYFWWTQSIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRHHQGPKRAESQDRKIFRGLICCYGPTNNMPTDQLEMMVOLCGASVVKELSSFTL 1800
Db 1741 VNGRHHQGPKRAESQDRKIFRGLICCYGPTNNMPTDQLEMMVOLCGASVVKELSSFTL 1800
QY 1801 GTGVHPVIVVQPDANTEDNGFHAIGQMEAPVVTREWVLDVSVALYQCOELDTYLIPQIPH 1860
Db 1801 GTGVHPVIVVQPDANTEDNGFHAIGQMEAPVVTREWVLDVSVALYQCOELDTYLIPQIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863

RESULT 3

US-09-734-672-6
; Sequence 6, Application US/09734672
; Publication No. US20020183268A1

GENERAL INFORMATION:

APPLICANT: Murphy, Patricia D.
Allen, Antonette C.
Alvares, Christopher P.
Critz, Brenda S.
Olson, Sheri J.
Schelter, Denise B.
Zeng, Bin

TITLE OF INVENTION: Coding Sequences of the Human

BCRAL Gene

NUMBER OF SEQUENCES: 72

CORRESPONDENCE ADDRESS:

ADDRESSEE: Morgan Lewis & Bockius LLP
STREET: 1111 Pennsylvania Ave., N.W.
City: Washington
STATE: District of Columbia
COUNTRY: USA
ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/734,672
FILING DATE: 03-Dec-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/966,436
FILING DATE: 07-No. US20020183268A1-97
APPLICATION NUMBER: US 08/598,591
FILING DATE: 12-Feb-96
ATTORNEY/AGENT INFORMATION:
NAME: Michael S. Tuscan
REGISTRATION NUMBER: 43,210
REFERENCE/DOCKET NUMBER: 44921-5055-02-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-739-3000
TELEFAX: 202-739-3001
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863 amino acids
TYPE: amino acid
STRANDEDNESS: No. US20020183268A1 Relevant
TOPOLOGY: No. US20020183268A1 Relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
STRAIN: BCRAL
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 17
MAP POSITION: 17q21
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-734-672-6

Query Match 99.9%; Score 9635; DB 9; Length 1863;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1860; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MDLSALRVEEVQNVINAMOKILECPICLLELKEPVSTKCDHIFCFCKMLKLLNKKGPQ 60
Db 1 MDLSALRVEEVQNVINAMOKILECPICLLELKEPVSTKCDHIFCFCKMLKLLNKKGPQ 60
QY 61 CPLCKNDITKRSLOESTRFSQVLELLKIIICAFQDITGLYANSYNFAKKENNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRFSQVLELLKIIICAFQDITGLYANSYNFAKKENNSPEHLKD 120
QY 121 EYSIIQSMGYNRNKRLLQSEPNPSLOETSLVSQLSNLGTVRTLTORKIORIQPKTSYVI 180
Db 121 EYSIIQSMGYNRNKRLLQSEPNPSLOETSLVSQLSNLGTVRTLTORKIORIQPKTSYVI 180
QY 181 ELGSDSSEDTVKNATYCSVGQDQLQITPQGTTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDTVKNATYCSVGQDQLQITPQGTTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNNDLNTTEKRAAERHPEKYQSSVSNLHVPCGTNTNTHASSLQHNSSLLLTKDRMNYE 300
Db 241 PSNNDLNTTEKRAAERHPEKYQSSVSNLHVPCGTNTNTHASSLQHNSSLLLTKDRMNYE 300
QY 301 KAFCNKSQKPLARSQNRWAGSKETCNDRTPTTEKVDLNADPLCERKEWNKQKLPC 360
Db 301 KAFCNKSQKPLARSQNRWAGSKETCNDRTPTTEKVDLNADPLCERKEWNKQKLPC 360
QY 361 SENPRDTEVPWITLNSSIQKWNFSDLELLGSDSDHDSSESNKAVADVLDVLINEVD 420
Db 361 SENPRDTEVPWITLNSSIQKWNFSDLELLGSDSDHDSSESNKAVADVLDVLINEVD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKKASLPNLSHVTE 480
Db 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKKASLPNLSHVTE 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRRTPSGLHPDDFTKKADLAVOKTPPEMNOGTNOTE 540
Db 481 LIIGAFVTEPQIIQERPLTNKLRKRRTPSGLHPDDFTKKADLAVOKTPPEMNOGTNOTE 540

[illegible]

Db	1731	GSTSVS-DGSLSVSTS-----LRKESVSESTIS-----LGGSQ	1767
Qy	1235	NNIPQSQRHRTVATECLSKNTEENLLSKNLNCSNOVILAKAQEHLSSETKCSAS	1294
Db	1764	SMSDSVSTSDSSLSVSTSORSES-VSESDSLSDSKSTSGSTSTSGSLSTSTSLSGS	1822
Qy	1295	LFSSQCSLEL---LTANTNTQDP-----FLIGSSKOMRHQSESQGVGLSDKELVSD	1344
Db	1823	ESVSESSLSDSISMSDSTSTSDSLSGSISLSGSTSLSTSDSLSDSKSLSSSQMSGS	1882
Qy	1345	EERGTLGLENQEEQS---MDSNLGEAAGCSSEFVSDECSGLSSOSDILITQOQRTWQ	1401
Db	1883	ESTSTSVSDSOSSTNSQFDSMSISASESDSMSTSDSSISGSNSTSTSLSTS---DSMS	1940
Qy	1402	HNLIKLOQEMAELEAVLPHQSGPSNSYPIISDSASLEDLRNPEOSTSEKAVLTQKSS	1461
Db	1941	GSV-----SVSTSTSLSDISGSI SVSDSSSTSTSESLSDMAQSQTSTASGSLSTSI	1995
Qy	1462	EYPISONPEGLSADKFEVSAOSST-----KNKEPQVERSSPKCPSLDDRWMY	1510
Db	1996	SLMSASAGTLTQSQTSVSTSLSTSDSISDSTISISGQSQAVESESTSDSTSIDSESL	2055
Qy	1511	HSCSGSLNARNYPQEEILKVVDFEQQLEESGPHDLTFTSYLPQDLGCTPYLESGISL	1570
Db	2056	-STGSSSTSTSTSESLSTNSGSTSVSES-----LSTSGGSTSVSDSSSTSSLSLST	2109
Qy	1571	FSDDPESDPSEDRAPEARV-----GNIPSTSTALKVPQLKVAESAQSPAAAHPTD--	1621
Db	2110	SGSTSVSDSTSMSESNSASISMSQSISGSTSGSTISISSELSMSGSTHNSVSDSDSI	2169
Qy	1622	-----TAGY-----NAM-----EESVSRKPELTASTERV	1646
Db	2170	STNSGSGMSNSIRHFTSLSTGLMSLSDSNSMDSDSVSISASEMSASMSDSVSMDSST	2229
Qy	1647	NKRSMWVSGLTPEEFMLVYFARKH-----HITLTLNLTET-----	1684
Db	2230	SSSMNSMSMTSESNST-----HPSMSMTSQTHFTSTSISTSESIAPNTNESQSTL	2282
Qy	1685	--THVVMKTDAE	1694
Db	2283	SATSVSSKHDAE	2294
RESULT 5			
US-09-815-242-12996			
; Sequence 12996, Application US/09815242			
; Patent No. US20020061569A1			
; GENERAL INFORMATION:			
; APPLICANT: Haselbeck, Robert			
; APPLICANT: Ohlsen, Kari L.			
; APPLICANT: Zyskind, Judith W.			
; APPLICANT: Wall, Daniel			
; APPLICANT: Trawick, John D.			
; APPLICANT: Carr, Grant J.			
; APPLICANT: Yamamoto, Robert T.			
; APPLICANT: Xu, H. Howard			
; TITLE OF INVENTION: Identification of Essential Genes in			
; FILE OF INVENTION: Prokaryotes			
; FILE REFERENCE: ELITRA.011A			
; CURRENT APPLICATION NUMBER: US/09/815.242			
; CURRENT FILING DATE: 2001-03-21			
; PRIOR APPLICATION NUMBER: 60/191,078			
; PRIOR FILING DATE: 2000-03-21			
; PRIOR APPLICATION NUMBER: 60/206,848			
; PRIOR FILING DATE: 2000-05-23			
; PRIOR APPLICATION NUMBER: 60/207,727			
; PRIOR FILING DATE: 2000-05-26			
; PRIOR APPLICATION NUMBER: 60/242,578			
; PRIOR FILING DATE: 2000-10-23			
; PRIOR APPLICATION NUMBER: 60/253,625			
; PRIOR FILING DATE: 2000-11-27			
; PRIOR APPLICATION NUMBER: 60/257,931			
; PRIOR FILING DATE: 2000-12-22			

;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 12996
;; LENGTH: 6281
;; TYPE: PRT
;; ORGANISM: Staphylococcus aureus
US-09-815-242-12996

Query Match 3.5%; Score 341; DB 10; Length 6281;
Best Local Similarity 18.3%; Pred. No. 1.1e-09;
Matches 335; Conservative 317; Mismatches 776; Indels 404; Gaps 79;

QY 7 RVEEONVINAMKILEPICLIEKIPKSTKCDHIFCKMCLKLLNKKGPSCQCLCKN 66
DB 2848 QVFAAMNOVNATKAALNGTQNLKAKQHANTAID-----GLSHLTNAQK 2891

QY 67 DITKRSLOESTRESQVLEELKIKICAFOLDTGLE-----YANSYNFAKKNNSP 115
DB 2892 EALKQLVQOSTTVAEAGGNEQK---ANNVDAAMDKLRQSTADNATTKQNGNYTDASQNK 2948

QY 116 EHLKDEVSIIQSGYRNRKRLLOSPEPNSLOETSLVSQSLNLTGRV-TLRTKQRIQ 174
DB 2949 DAYNNAVTTAQG-----IIDQTSPTLPTVINQAAGQVSTTKNALNGNENLEA 2998

QY 175 KTSVYIELGSDSSEDTVNKATYCSVGQ-----ELQITPQGTREDEISLDSAKKAAC 226
DB 2999 KQASOSLG---SLDLNNAKQTVTDQINGAHTVDEANOIKQNAONLNTAMGNLKOATA 3055

QY 227 EFSETDVTNTEHROPNSNDLNTTEKRAAEHPEKYOGSSVSNLHVPPCGTNTHASSLQHE 286
DB 3056 DKDATRAT-----VNETDAQKQQAAYNTAVTNAENIISKANGGNATQAEVEQA 3104

QY 287 NSULLITKDMNVKEAFCKNSKQPLGARSQHNHWAGSKETCNDRTTPSTEKKVYDLNADP 346
DB 3105 IKOVNAKQALN-----GNANVQH-----AKDEATALINSNDLNQ-- 3140

QY 347 LCBKRWKOKLPCSNPRDTEVPWITLNSIQKNEWFSRDELLGSDSDHSGESEN 406
DB 3141 --AQKALKQOVQONATVAGVNV-----KQTAQELNNAMTQLKQ--GIADKEQTADGN 3191

QY 407 AKVADVLVNEVDEYSGSEKIDLLASDPHEALICKSERVSHKSVESNIEDKIFGKTYR 466
DB 3192 FVNADP-----DKONAYNOAVAKAEALISATPDVVVTPSEITAALNKVTVQAKNDLNGNTNL 3247

QY 467 KKA-----SLPNLSHV-----TENLIIGAFVTEPQIIQRPILT-----NKLK 504
DB 3248 ATAKQNVQHAIDQLPNLNAQRDEYSKQITQATLVPNVNAIQQAATTLNDAMTQLKOGIA 3307

QY 505 ---KRRPTSGLHPEDFIKKA--DLAVOKTPEMINOCTNOT-----EONGQVMNITNSG- 552
DB 3308 NKAQIKGSYENYHDADTDKQATYDNATYKAEELKQNTNTPMDPNTIQOALTQVNDTNQAL 3367

QY 553 HENKTGKDSIQNEKNP-NPIESLE--KESAFKTKAE--PISSISINMELEL-NIHNKAP 606
DB 3368 NGNOKLADAKQDAKTTTLGTLDHLNDAKQALTTQVEQAPDIATVNVVKNQAQNLNNAWNT 3427

QY 607 KKNLRKSTRIHALELVVSRNLSPNCTELOIDSCSSSEBIKKKKYNOMPVRHSRLN 666
DB 3428 LNNALQDKTET-----LNSINFTD-----ADQAKKDAYTN-AVSHAEGI 3465

QY 667 QLMGKEPATGAKKSNPEOTSKR-----HSDTTPPELK-----LTNAPGSEFTKC 712
DB 3466 -----LSKANGSNASQTEVEQAMQVRNEAKQALNGNDNVQRAKDAKQVITNA----- 3513

QY 713 SNTSELKEFVNPSPREEKEKLETY----KVSNNAEADPKDMLSGERVQLQTERSVESS 768
DB 3514 ---NDLNQAKQDAL--KQOVDAQOTVANNVTIKQTAQDLNQAMTQLKQGIADKQDKANG 3568

QY 769 ISLVPGTD-----YGTQESISLLEVTLGKAKTEPNKVCVQCAAFENPKGLIHG-----CS 819
DB 3569 NFYNADTDKQANVNAVAHAEQIISGTPNANVDPQOVAQALQOVNAQKGLNGNHNLOVA 3628

QY 820 KDNRNTEGFKYPLGHEVNVHNSRETSTEMEESSELDQYQLNTFKVSKRQSFALFSPNGAE 879
DB 3629 KDNAN-TAIDQLP---NLNQPKTALKQDVSHAE---LVTGVNAIKQNDALNMMGTUK 3681

QY 880 EECATFSAHSGSLKKSPKVFEECEKEENQGNKSNIRKPVQTVNITAGFP-----VV 932
DB 3682 QQ-----IQANSQVPQSVDFT-QADQDKQQAAYNNAAN---QAAQIANGIPTVPLTPDVT 3731

QY 933 GQKDKPVDNAKCSIKGSRFCJSSQ--FRGNET--GLITPNKHGLLQNPYRPPLEPIKS 988
DB 3732 TQAVTTMNAQKDALNGDEKLAQAKQALANLDTLRLNQPQDRLNRQINQAQALATVEQ 3791

QY 989 FVTKTKCKNLEENFEHNSPEREMGEN--IPSTVSTISRNNT--RENFKKASSSN 1043
DB 3792 TKQNAONVNTAMSNLKGATANKDTVKASNYHDADADKQATYNAVQAEGIINTNT 3851

QY 1044 IN--EYGSSTNEVGSSINEIGSSDENIQABLGRNRGPKLNAMLRGLVLOPEVYKQSLPGS 1101
DB 3852 LNPDEITRALTOVTDAKN--GLNGEAKLATEKQNAKDAVSGMTHLN---DAQKQALKGQ 3905

QY 1102 NCKHPEIKQVEEYEVQTVNTDPSYLIISDNLEQPMGSSHASQVSETPDDLDDGEIKE 1161
DB 3906 IDQSPET--ATVNOVKQT-----ATSLDQAM--DQLSQAINDKAQTLADGNYLNA 3951

QY 1162 DTSFAENDIKE-----SSAVFSKSVQKGLSRSPSPFTHTHLAQGYRRGAKKLESSSENLS 1217
DB 3952 DPD-KONAYKQAVAKAEALLNKQSGTNEVQAQVESTINEVNA-----AKQALNGNDNIA 4004

QY 1218 S-----EDEELPCFQHLFLGKVNINIPQSQTRHSTHSTVATECLSKNTEENLLS-----LKNSLN 1268
DB 4005 NAKQAKQQLANLTHLNDAAKQSFESQITQ-APLVTDVTTINQAKOTLDHAMELLRNSVA 4063

QY 1269 DCSNQVILAKASQEHLSSETKCSALFSSQSELEDLTANTNTQDPFLIGSKQMRHOS 1328
DB 4064 D--NOTTL--ASEDYH-----DATAQRQNDYNOAVTAANNINOT 4099

QY 1329 ESOGVGLSD-----KELVSDDEERGTLGEBNNOEEOQSMDSNLGEAASGCESETS 1377
DB 4100 TSPMNPDDVNGATTQVNTKVALDGDENLAAKQKQANNRLDOLDHLNNAQKQQLQSQIT 4159

QY 1378 VSEDCGLSSQSDILTQQRDTMQLNLIKLOQEMAELEAVLEQHG-----SQPSNSYPS 1431
DB 4160 QSSDIAAVNGHKQ--TAESLNTAMGNLI---NAIADHQAV-EQGRGNFINADTKQFAYNT 4213

QY 1432 IISDSAL--EDLRNPEQSTSEKAVLTQKSSEYIPSONPEGLSAD-KFEVSAADSSTSK 1487
DB 4214 AVNEAAMINKQTQGNANQTEVEQAITYQVOTTL-----QALNGDHLNLOVAKTNATO- 4264

QY 1488 NKEPGVGRSPSKPCLDDRWYMHSCSGSLONRNYPQSOBELIKVVDVEE---OOLES 1542
DB 4265 -----AID-----ALTSUNDQKQALKDOVTAATVLTAVHOIQN 4299

QY 1543 GPHDLTETSYLPQDLLEGTPYLESIGLSFDDPESDPSEDRAPEASRVGNIPSSTSALKV 1602
DB 4300 A-NLNAQMHGLRQSIQDNNATKANSKYINEQPEQONYDQAVQA--NNIINEQTA--- 4353

QY 1603 POLKVAESAQSPAAAHITDTAGYNAMEEVSREKPELTASTERVNKR--MSWVVSGLTPE 1660
DB 4354 -----TLDNNAINAATTVNTTKAALHGDVQLQNDKHAKQTVSOLAH- 4396

QY 1661 EFMVYKFARKKHITLNLITETTHVMKTD 1692
DB 4397 -----LNNAQKH--MEDTLIDSETTAVTKQD 4421

RESULT 6
US-09-815-242-5635
; Sequence 5635, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.

```

; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELTRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5635
; LENGTH: 2368
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-5635

```

```

Query Match 3.48; Score 328.5; DB 10; Length 2368;
Best Local Similarity 18.28; Pred. No. 1.3e-09;
Matches 349; Conservative 306; Mismatches 720; Indels 547; Gaps 78;

```

```

QY 93 FOLDGTGLYANSYNFAKKNENSPHLKDEVSIIQSMGYRRA-----KRLQSE 141
DB 390 FINT--EIGNNGFG--ASLKAQDFKEVTLPGQVYVNNSLTTTFPNGNEDSTVLKNN 445
QY 142 PNPSLQTSLSVLNSLNGVTRTLTKRIOPQKT---SVYELGSSSEDTVN---KAT 195
DB 446 TVNYDONANKVFTSQGVTTARGTHTKEVLPDPKSLSKYKVNANIDTKNIDFNEKLT 505
QY 196 YCSVGDQELLOITPGCTRDELISLSAK--KAACEFSETDVTNTEHHQPSNNDLNTTEKRAA 254
DB 506 YRTASD-----IVINNAQPEVTLTADPFSVAVEMNKDALQQQVNSQVDNSHYTTAS---IA 558
QY 255 ERHPEKYQGSVSN---LHVFPK-----GNTTHASSLQHENSSLLLTDKRMNVEKA 302
DB 559 EYNKLLQQAADNLNEDANHVETANRASQAADGLVTKLQAALIDNQAALAEALDAKAEKV 618
QY 303 EFCNKSQPGGLARSOHNWAGSKETCNDRRTPSPTEKKVDLNADPLCERKE-----WNKQK 357
DB 619 TAAQGSKKV---TQDEVAALVTIKNNDKNAIAEINKQTAAQGVTTKONGIAVLDDQV 674
QY 358 LPCSENPRDTEVPWITLNSSIQKWNFWRSDELSCDSDSHDGESESNKAVADVLDVLN 417
DB 675 IPTPVKPAKODI-----IQAVT---TRKOI-----KKSNASLQDEKDVAN 713
QY 418 EYDEYSGSSEKIDLLASDPHICALICKSERVHSKSVESNIEDKIFGKYRKKASLPNLSHV 477
DB 714 D-----KIGKIEYKAIDDAATNQAQVEAIKTKAINDINQATAPATAKAAALEEDEV 767
QY 478 TENLI-----IGAFVTEPQIIQERPLTNKLRKRRTPSGLHPEDFIKKADLAVOKTPMIN 533
DB 768 VQAQIQADQAPLNPDTNNEEVAERINAAK-----VSGVK-----AIEATTTAQD 812
QY 534 OGNTQTEONGOMVNTNSGHENKTKGDSIQNEKNPNFIESLEKSAFKTKAEPISSISN 593
DB 813 LERVKNEEISKIENTDS---TQTKMDAYNEVK-----QAATARKTONATVSNATNE 861

```

```

QY 594 MELELNHNKAPKKNNRLRRKSSSTRHHALELVVSR-----NLS 632
DB 862 EVAEADAARAAQKQG-----LHDIQVVKSKQEVADTKSKVLDKINAIOQAQVK 911
QY 633 PPNCT-----ELQIDSCSSSEIEKKKKYNQMPVRHSR---NLQLMGKKEPATGA 678
DB 912 PAADTEVENAYNTRKQETQNSNASTTEE-KQAAYTELDTKKQEARTNLDANTNSDVTTA 970
QY 679 KKS-----NKPNEOTSKRHSDTPELKLNTNPGSFTKCSNTSELKRFVNPSLPREEK-- 731
DB 971 KNGIAAINQVQAATTK--SDAKAEIA-----QKASERKTAIEAMNDSTTEEQAA 1020
QY 732 EEKLETYKVSNNADPKDMLSGSERVLQTERSVESSISLVPDGYGTQESISLLE---- 787
DB 1021 KDKVDQAVVTANAD--IDNAAANTVDVNAKTTNEATIAAITPDANVKPTAKQAIADKVA 1078
QY 788 ----VSTLGRKAKTEPNKCVSQCAAFENPKGLIHGCSKNRNDTEGFKYPLGHEVNHRET 843
DB 1079 QETAIDANNGATTEEKAQKQVQTEK-----TTADTAID-----GAHTNAEVEA 1123
QY 844 SIEMEESELDAOYLQNTFKVSKROSEFALFSPG-----NAERBCATFSAHSGSLK 893
DB 1124 AKNAEIAKIEAIPATTTKNAKQAIATKANERKTAIAQTODITAEETAAANANVDNAV 1183
QY 894 KQSPKVTFECEKEENQK--NESNIKPV-OTV-----NITAGFPVVGQKDKP 938
DB 1184 QANNIEAANSQNDVDQAKITGEASIDQVTPVNKKATAVTDKANNITA-----ATDDNG 1238
QY 939 VDNK-----CSIKGGR-----FCLSSQFRNETGLITPNKHGL----- 973
DB 1239 VDTAKDAGKNSIOSTQPATAVKSNKNDVDQAVTTQQAIDNTTGAETEEKNAARDVLK 1298
QY 974 -----LONPYRIPPLPIK-----SFVKTCKKNLLNEEHEHSM----- 1008
DB 1299 AKEKAYODILNAQTNDVTQIKQOAVADVQGITADTTIKQVAKDELATKAREQKALIAQT 1358
QY 1009 -----SPERE-----MGNEIPSTVSTISRNIRENVFKEASSNINEVGSSTN- 1052
DB 1359 ADATTEKEQANQOQVDAELTQGNQNIENAOISDDVNTAKDNAIOA-----IDPIQASTDV 1413
QY 1053 -----EVGSSINEIGSSDENIOAELGRNRP-----KUNAMRLG-----V 1088
DB 1414 KTNARABELLTEMQNKITEILNNNETTNEEKGNDIGPRAAYEEGLNNINAATTTGDVTTA 1473
QY 1089 LQPEVYK-QSLPGSNCKHPEIK-----KOEYEVVQTVNTDFSP 1126
DB 1474 KDTAVQVQQLHNPVKRPKPKACTALDQAAADKTKQIEQTPNASQOEINDAKQEVDTLQ 1533
QY 1127 YLISDNLEQPMGSSHASQVCSQVSETPDDLLDDGE-----IKEDTSAENDIKESSAVFSKV 1181
DB 1534 --AKTNIDQ-----SSTDEYVDNAVKEGKAKINAVKTFSEYKKDALKAKIEAAYNAKV 1583
QY 1182 QKGLSRSPSPFTHLAAQGYRRGAKKLESEENL-----SSEDEELPCFQHL----- 1229
DB 1584 TEADNSNAS---TSSEIAEAKQKLAELKQATADQNVNQAATSKDDIEVQIHNDLDINDYTI 1640
QY 1230 -----LFG-----KVNIPQSQTR---HSTVATECLSKNTEENLLSKSLND- 1269
DB 1641 PTGKESATDLYAYADQKRNKNSADTNAQDEKQQAQKQVQNVQTALESINNGVDNGD 1700
QY 1270 -----CSNQVILAKASQEHLSSEETKCSASLFSQCSQSEDL 1306
DB 1701 VDDALTQGAADIAIOVDATVKPKANQVIDAKA-----EETK-----ESIDQSDQL 1746
QY 1307 TANTQDQPLIGSSKQMRHOSQSGVGLSKDELVSDEE--RGTGLE-----ENNOQE 1358
DB 1747 TABEKEALAMI---KOITDQAKQ---GITDATTTAEVERAKAQGLEAFDNQIDSTEK 1800
QY 1359 QSMDSNLGEARASCESETSVSEDCSGLSSQSDILTTQORTMOHNL-----IKLQ 1408
DB 1801 KALE-----ELETALDQTEAGVNVQAD-ATTEKEAFTNALEDILSKATEDISQ 1849
QY 1409 QEMAELEAV----LEQHGQSQPSNSYPSIISDSSALEDLRNPEQSTSEKAVLTOKSEY 1464

```

Db 1850 TTNAETATVKSALQOLKAQRINP-----VVKNALEAIREVNVKQIEIINKNADASAKE 1905
Qy 1465 ISONPEGLSADKFEVSADSSKNKE-----PGVERSSPSKPSLDDRWYHMSCG 1515
Db 1906 IARTDLGRYDFRFBADKLD-KTQNTNEVABLQWVTIPAEIAIYPQNDPNAND-----TNSG 1959
Qy 1516 S-----LQNRNPSQBELIKVVDV----- 1534
Db 1960 SONNDATANSANATPENTGQPNVTESTDNADATSSITTNNQNDAAATGETTATSANSSA 2019
Qy 1535 -----EEQOLESSEGHDTETSYLPRQLEGPYIESGLISLSEDDPESDPAESAR 1589
Db 2020 TDDANDKFOANNSSADTSTNSPTMDNDVTSKPEVESTNNGTTRDKPATEADNATPAESAT 2079
Qy 1590 VGNIPSSTSALKVPOLKVAESAQSPAAAHHTTDTACYNAMESV-----SREKPELTAS 1642
Db 2080 NNNSTTTATNENAP-----TESTATAPTASTGAESSADSKONASVNDSKONAEVNN 2132
Qy 1643 TE 1644
Db 2133 AE 2134

RESULT 7

US-09-815-242-12389
; Sequence 12389, Application us/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815.242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12389
; LENGTH: 2368
; TYPE: prt
; ORGANISM: Staphylococcus aureus
US-09-815-242-12389

Query Match 3.4%; Score 328.5; DB 10; Length 2368;
Best Local Similarity 18.2%; Pred. No. 1.3e-09;
Matches 349; Conservative 306; Mismatches 720; Indels 547; Gaps 78;

Qy 93 FOLDNGLEVANSYFAKKENNSPEHLKDEVSIIQSMGYNRA-----KRLQSE 141
Db 390 FTINT--EIGNNGNFG--ASLKADQFKYEVLPOGVTVVNSLTTFFPNGNEDSTVLKNN 445
Qy 142 PENPSLQETSLSVQLSNLGTVRTLRTKRIQPKQT---SVVIELGSDSDSEDTVN---KAT 195

Db 446 TVNYDONANKVFTTSGQVTTARGTHTKEVLPDKSLKLSYKVNANIDTPKNIDENKLT 505
Qy 196 YCSVGDQELLQITPOGTDEISLDSAK-KACEFSETDVTNTEHHQPSNNDLNTTEKRAA 254
Db 506 YRTASD-----IVINNAQPEVLTITADPFVSIVEMKMDALQOOVNSQVDSNHYTTAS--IA 558
Qy 255 ERHPKPYGSSVSN---LHVEPC-----GTNTHASSLQHENSLLLTKDRMNVEKA 302
Db 559 EYNKLIKQADNINLEDANHVETANRASQAADGLVTKLQAAALIDNQAIAELDAKAQEKV 618
Qy 303 EFCNKSKPGIARSOHNRWAGSKETCNDRRTPSTPEKKYVDLNADPLCERKE-----WNKOK 357
Db 619 TAAQOSKKV---TQDEVAALVTKINNDKNAIAEINKQTTAQGVVTEKDNGLAVLDQDV 674
Qy 358 LPCSENPROTEVDVPIITLNSSITQKYNWFSDLSGDSHSDGSESNKAVADVLDVLN 417
Db 675 TPTVTKPOAKODI-----IQAVT---TRKQQT-----KKSNASLQDEKDVAN 713
Qy 418 EVDEYSGSSEKIDLLADPHALICKSERVHSKSVESNIEDKIFGKYRKKASPLNLSHV 477
Db 714 D-----KIGKIETKAIKOIDAATTNAQVEAIKTAINDINQOTAPATTAKAAALEEDEV 767
Qy 478 TENLI-----IGAFVTEPQIIQERPLTNKLRKRRTPSGLHPEDFIKKADLAVQKTPEMIN 533
Db 768 VQAQIDQAPLNPDTTNEEVAEAIERINAAK-----VSGVK-----AIEATTTAQD 812
Qy 534 QGTNQTQEQGVMMNITNSGHENKTKGDSIQNEKNPNPESLEKESAFKTKAEPISSSISN 593
Db 813 LERVKNEEISKIENITDS---TQTKMDAYNEVK-----QAATARKQNTATVSNATNE 861
Qy 594 MELELNIHNSKAPKKNRLRRKSTRHIALELVVSR-----NLS 632
Db 862 EVAEADAVERAAQOG-----LHDIVVKSQEVADTKSVKVLKINAIQTOAKVK 911
Qy 633 PPNCT-----ELQIDSCSSSEIEKKKKYNQMPVHRSR---NLQMEGKEPATGA 678
Db 912 PAADTEVENAYNTRKQEIQNSNASTTEB-KQAAAYELDTKKQEARTNLDAANTNSDVTTA 970
Qy 679 KKS-----NKPNEQTSKRDSDTFFPELKLTNAPGFTKCSNTSELKEFVNPSLPREEK-- 731
Db 971 KNGIAAINOVQAAATTK--SDAKAEIA-----QKASERKTAIEAMNDSTTEEOQAA 1020
Qy 732 BEKLETVKVSNNAEDPKDMLSGERVLTQERSVESSSISLVPDGTDTGYQTESISLE---- 787
Db 1021 KDKVDQAVVTANAD--IDNAAANTVDNAKTTNEATIAIITPDANVKPTAKQAADKQVA 1078
Qy 788 ---VSTLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNRRNDTEGFKYPLGHEVNHRSRET 843
Db 1079 QETAIDANNGATTEEKAQKQVQTEK-----TTADTAID-----GAHTNAEVEA 1123
Qy 844 SIEMEESELDAQYLQNTFKVSKRQSFALFSPG-----NAAEECATFSAHSGSLK 893
Db 1124 AKNAEIAKIEAIOPATTTKDNAKQAIAATKANERKTAIAOTQODITAEETAAANANVDNAT 1183
Qy 894 KOSPKVTECEQKEENOCK--NESNIKPV-OTV-----NITAGFPVVGQKDKP 938
Db 1184 QANNIEAANSQNDVDQAKTTGEASIDQVTPVVKKATAVTDAKNNITA-----ATDDNG 1238
Qy 939 VDNAK-----CSIKGGSR-----FCLSSQFRNETGLITPNKHGL----- 973
Db 1239 VDTAKDAGKNSIQSTQPATAVKSNNAKNDVDQAVTTQNCQIDNTTGTATTEEKNAAKDLVLK 1298
Qy 974 -----LQNPYRIPPLFPPIK-----SFVKTKCKKNLLEENFEESHM----- 1008
Db 1299 AKEKAYQDILNAQTNDVTDQKQAVADVQGITADTTIKDKVAKDELATKAREQKALIAQT 1358
Qy 1009 ---SPERE-----MGNENIPSTVSTISNNRENKVFKEASSNSINEVGSSTN- 1052
Db 1359 ADATTEEQANQAVDAELTQCNQNIENAQSIDDVNTAKDNAIQ-----IDPQASTDV 1413
Qy 1053 -----EVGSINIEIGSSDENIQAEIQLGNRNP-----KLNAMLRLG-----V 1088

Db	697	RISRKK-----KPSPPKVLPEPNQKQEKTEKEEKTWNGR	732
Qy	922	TVNITAGFPVVGQKDPVNAKCSIKGSRFCLSQFRGNETGLITPNKHGLLQNPYRIP	981
Db	733	TLRRSPRISRTAKVAEIRDQADKKRGG-----EDEVEEESTALQTKDKKEIL	782
Qy	982	PLPIKSFVTKCKKLLBENTFEESMSPE---REMGNENIPSTVTSIRNIRENVKE	1038
Db	783	-----KKSEKDTNSKVSKVRPKGVRWTGR-----TGRWKYSNDESE	822
Qy	1039	ASSNINEVGSSTNEVSGSSINEIGSSDENIQAEILGRNRGPKL-----NAMLRLGVLOP	1091
Db	823	GSSEKSSAASEEKESEEAAILADDDPECKKGLPNHPELILLDCSDSGYHTACLRP	882
Qy	1092	EYVKQSLPGSN-----CKHPEIKQYEEVQVTVNTDPSYLISDNLEQPM---GSSHAS	1143
Db	883	PL--MIIPGWEFPCPQH--KLLCEKLEFEOQLDLDVALKKRAERRKERLRYVVGISIE	939
Qy	1144	QVCSETPDLLDDGEIKEDTSA-----ENDIKE	1172
Db	940	IIPQEPDFSEDEEKKDSKSKANLERRSTRPRKICISYRFEDEDAIDEALDDIKE	999
Qy	1173	SSAVFSKSVQGBLSRSPFTHLQAGYR--RGAKKLESSEENLSSDEELPCFQHLFF	1231
Db	1000	ADG-----GGVGRGKISIT--GHRGKDISTLDEERKENRKPQAAAAARRKR	1047
Qy	1232	GKVNPNPSQSTRHSTVATECLSKNTEENLLSKNLDCNSQVILAKASQDHLSSETKC	1291
Db	1048	RLNLDSDSN-----LDEESESDEFKISDGSQD-----EFVYSDENPD	1086
Qy	1292	SASLFSQCSLEDLTANTNTODPFLIGSSKOMRHOSESQGVG-----LSDKELVSDDEE	1346
Db	1087	ESE-----EDPNSNDSOTDTC--SRLLRHPSPRPMQSRRLRKPYPKKYSDD-	1134
Qy	1347	RGTGLENNQEOGSDSNLGEAASCESETSVSDECSGLSSQSODILTTQORDTHQHLNK	1406
Db	1135	-----EEEESEENSRD-----GESDFSDDFS---DDFVETRRRRSRRN---	1169
Qy	1407	LOQEMAELEAVLRHCGSQPSNSYPSTISDSSALEDLRNPEQSTSEKAVLTSOKSEYPI	1466
Db	1170	QKQRYNKEDSDGSQSKSLR-----RKEIRRVHKKRLLSSESESYLS	1214
Qy	1467	QNPEG-----LSADKFEVSADSTSKNKEPQGVERSPKCPSLDDRWMH-----	1511
Db	1215	KNSDEDLAKESKRSVRKRGSTDEVSEADEEBEEEGKPS-----RKLRIETDEE	1267
Qy	1512	-SCSGSLQNRNYPQOEELIKVVVDVEEQOLESQPHDLTETSYLPRQDLE-----GTPYLE	1565
Db	1268	ESCDNAHGANOAPARDQSRPLPSQESTKK--PYRIESDE---EEDFENVKGVGSPLDY	1322
Qy	1566	SGISLFSDDPESDPSDRAPESARGVNPISPTSAUKVPOLKVAESAQSPAAAHYTDTAG-	1624
Db	1323	SLVDLSTNGQS-----PGKAIENLICKPTEKSQTPK-----DNSTASASLASNGTSGG	1371
Qy	1625	--YNAMEESVSRKPKELTASTERN	1647
Db	1372	QEAGAPEE-----EEDELLRTVLVD	1392

RESULT 10

```

US-09-815-242-5916
; Sequence 5916, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of

```



```
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34248
; LENGTH: 2665
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL034555.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 9.3
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 12
; OTHER INFORMATION: EST HUMAN HIT: A0117052.1, EVALUATION 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P08640, EVALUATION 3.00e-10
US-09-864-761-34248
```

Query Match 3.2%; Score 308; DB 10; Length 2665;
Best Local Similarity 19.0%; Pred. No. 2e-08;
Matches 375; Conservative 275; Mismatches 724; Indels 598; Gaps 89;

```
Db 424 RKEILKRESKKIKLRLNTVASPKDQCEL-----ASISVSGSGSRPSSDLOARLGLAGESV 479
QY 200 GDOELLQITPQGTREI-----SLDSAKKAACEFSETDVTNTEHHOPSNNDLNT 248
Db 480 ENQEVOSKKPIPSKPOLKQLOVLDDOQPEREDVRKNKYCSLRD-ETPERKSGQEKSHSVNT 538
QY 249 TEKRAAE-RHPEKY-----QGSVSNLHVPCGTNTHAS-----SIQHE----- 286
Db 539 EEKIGIDIDHTQSYRKQMEQSRKQOMEMELAKSEKFGSPKDVDEYERSLVHEVGKPP 598
QY 287 ----NSSLLLTDRMNVKAEFCNK-----SKOPGLARS-OHNRW----- 321
Db 599 QDVTDDSPSKKRMHDVDFDICTKRERNYRSRQISEDSEGTGGSPSVRHGSHFDEDDP 658
QY 322 AGSKETCNDRTPTSTTEKKVDLADPLCERKE---WNKQKLPCESENPRDTEDPWITLNS 378
Db 659 IGSPLLVSXGSPKVDKVL-LPYSNITVREESLKFNPYDSSRRREQMADMAKIKLSVLNSE 717
QY 379 IOKVNEWFSRDELIG-----SDDSHGSESNK 408
Db 718 -DELNRWDSQMKQDAGRFVSPNSIIKRDSLKRKSVRDLPEGCEVPDSDDEGEHKSHSP 776
QY 409 VADVLDVLNEVDEYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKK 468
Db 777 RASAL-----YESSRSLFLLREDKLRERDERL-SSSLERN-----KPY--- 815
QY 469 ASLPNLSHVTENLIIGAFVTEPOIIOERPLTNKLRKRRTSGLHPEDFIKKADLAVQKT 528
Db 816 -----SFALDKTI---TPTKALLERAKSLSSREENW----- 845
QY 529 PEMINOGTNQTONGQVMNITNSGHENKTGDSIQNEKNPNPIES-LEKESAFKTKAEP 587
Db 846 -----SFLDWDSRFANFRN--NKDEKVDSA-----PRPIPSWYMKKKIRTDSE-- 888
QY 588 SSSISNWELELNHNSKAPKNRLRKRKSTRHIALELVVSNLSPNCTELQ-----I 641
Db 889 ----GKMDKEDKHEEQEQL-----FASRFLHS-----SIFQDSKRLQHLERKEE 934
QY 642 DSCSSEIEIKKKYN-----OMPV--RHSRLQL--MEGKEPATGAKSKNPNQ 687
Db 935 DSDFTSGRIYQKOTSEGANSTWDSIOEPVVLFSRPMELTRMOQKE-----KEDQKPK- 989
QY 688 TSKRHDSDFPELKLITNAPGSETKTSNTSELKEFVNPSPREEKEKLET-----VK 739
Db 990 VEKQEDTENHPK-----TPESAPENKQSELKTPPSVGPPSVT 1026
QY 740 VSNNAEDPKDL-MLSGERVLTQERSVESSISLVPGTDYGTQOESISLLEVLSTGLAKTTP 798
Db 1027 VVTLESAPSALEKTTGDKTVEAPLVEEKTVE--PAT-----VSEAKPASEP 1072
QY 799 NKC-VSOCAAFENPKGLIHGCSKDNNDTEGKYPLGHEVNHRSRE-----TS 844
Db 1073 APAPVEQLQVLDLPPGA-----DPDKEAAMPAGVEEGSGDQPPYLDKAPPTPGAS 1124
QY 845 IMEESSELDQYLQNTFKVSKRQSFALFNSPNAEBCATFSAHSGSLKK-----QSPK 898
Db 1125 FSOAESNVDP-PSDTPQLSKPAQKSEANEKAEKPDATADAEPDANKAEAPESQPP 1183
QY 899 VTPECB-----QKEENQCKNESNTKPVOTVNITAGFPVVQKDKPVDNAKC-----SIKG 948
Db 1184 ASDELEVDPPVAAKDKKPNKSRKSTPVQAAAVSIVEKPVTRKSIDREKLRKRSNPRG 1243
QY 949 GSRFCLSSQFRGNE-TGLITPNKHGLLQNPYRIPPL-----FPI 986
Db 1244 EAQKLELMEAEKTPRTASKNSAADLHEPEPSLPSLRTRRNRVRSVYATMGDHNRSVP 1303
QY 987 KSFV-KTKCKKNLLEENFEHNSMP-----EREMGNE----- 1017
Db 1304 KEPEVQPRVTRKLELQEAFAVPTTPRRGPPKTRRRADDEENEAKEPAETLKPPG 1363
QY 1018 -NIPSTVSTIS-----RNNIRENVKEASSNINE---VGSSTNEVGSINIGSS 1064
Db 1364 WRSPRSQKTAAGGGPOGKKGKPKVDATRPETATTEVGPIGVKESMEPKAAEEAGSE 1423
```


QY 851 ELDA-----QYLONTFKV--SKRQFALFSPNPNABEECATFSA-----HSGSLKKQS- 896
Db 831 LANAKQPKQOLANLHLNDAQKQSFESQIQAPLVTDVTTINQKAQALDHAMELLRNSI 890
QY 897 --PKVTFECE-----QKEENQ--CKNESNIKPQVOTNITAGFPVGVQKDKPVDN 941
Db 891 ADMQATLASEDYHDATAQKQNDYNQAVTAKNLIIN-----QTSPTWNPDEVNKRATTQVNN 946
QY 942 AKSIRGGRFCLSSQFRNETG---LITPNKHGLLQNPYRIPPLFPFKISFVKTKCKKN 997
Db 947 TKVALDGDENLAQAQKQANRLNQLDHLNNAQKQQLQSLQIAQSSDIAAVNGHKQTAESLN 1006
QY 998 LLEEN-----FEHSMSPERBMGN-----ENIPSTVSTISRNIRENVFKE-ASSSNTINEVG 1048
Db 1007 TAMGNLINATAIDHOAEVQOR--GNFINADTDKQATYATVNEAEAMINKOTQGNANQPEVE 1064
QY 1049 SSTNEVGSSINEIGSSDENITQAEFLGNRGPKNLMLRLGLVQPEVYK-----1095
Db 1065 QAITKVQTTLOAL--NGDHLQV--AKTNATQAI DALTSNDPQKALTQDQVTAATLVTAHV 1122
QY 1096 -----QSLPG-----SNCKH-----PEIKKQYEEVVQ-----1118
Db 1123 QIEQNANTLQAMHGLRESIQDQNAATKANSKYINEDQPE--QONYDQVQAANSIINEQT 1180
QY 1119 -----TYNTDFSPYLISNLEQPMGSSHASQVSETP--L-----DDLDD 1156
Db 1181 ATLDDNAINQAATVTNT--TKAALHGDVGLQNDKDKAKQTVSGLAYLNNAQKHMEDTLID 1238
QY 1157 GE-----IKED-----TSPAENDIKESSAVE-----1177
Db 1239 SETTRIVAKQDLTEAQAALQDLQMLNTLOOSTADKDATRASSAYVNAEPNKKQAVDEAVONAE 1298
QY 1178 -----SKSVOKGELSRSPSPFTH--HLAQGYRRGAKKLESSEENLSEDEELPCFQHL 1229
Db 1299 SIITAGLNNPTINKGNVSSATQAVTTSKNGLDGVRLAQDKQTAGNSLNHLDLQITPAQQA 1358
QY 1230 LFGKVNIPQSOTRSTHVTATECLSKNTEENLLSKNSLDCSNQVILAKASOEHLHUSEET 1289
Db 1359 LENQINN-----ATTRDVAEITIAQAQALNEAMKALKESIKDQOTEASSKFINEQDQAOKA 1415
QY 1290 KCSASLFSQCSLELDTANTQDPLFLGSSKQMRHOSQGV--GLSDKELVSDDEERG 1348
Db 1416 -----YTOAVQAHKDLI--NKTDTPLVKSVIDQATQAVNDAKNLHGDQKLAQDKQRA 1467
QY 1349 TGL-----BENQOEQMSDNLGEAASGCESETSVSDCGSLSSQSDIILTQOQRTMOHN 1403
Db 1468 TETLNNLSNLTPOQALENQINNAATRGVAQKLTE-AQALNQAMEAL----RNSIQ-- 1520
QY 1404 LIKLOQEMAELEAVLBQHGQSQNSYPSIISDSSALEDLRNPQESTSEKAVLTSQKSSEY 1463
Db 1521 ----DOOQTEAGSKFINKEDPKQKDAYQAQVQAHAKDL--INQTSNPTLDKA-----1564
QY 1464 PISONPEGLSADKFEVSADSSTSKNKEPVERSSPKSPCLSDRWYMHSCSGSLQNRNYP 1523
Db 1565 QVEQLTOGVNQAKDNLHGQKLA-----DDQHAVTDLNQLNSLNNP 1606
QY 1524 SOBELIKVVDVEEQOLESGPHD-----LTETSYLP-----RODLEGPTPYESGISLFS 1572
Db 1607 QROAL-----ESQINNAATRDVAQKLAQALQDQAMQALRNSIQDQOQTESGSKFEIN 1659
QY 1573 DD-PESDPSDRAPEARSVGNIPSTSSALKVPQLKVAESAQAAPAAHTTDTAGYNAMES 1631
Db 1660 EDKPKQD-----AYQAA-----VQAHKOLINQTNPTLIDKQSQVEQLTOA 1698
QY 1632 VSREKPELTASTERNVKNRSMVVSGLTPEEFMLVYKFAKKHI--TLNLI-----TE 1682
Db 1699 VITAKNLHGDKLARDQOQAVTT-----VNALPNLNHAQQAALPDAINAAPT RTE 1749
QY 1683 ETHVVMKTDABEVCERTLK 1702
Db 1750 VAQHVTQATDELHAME-TLK 1768

RESULT 14
US-09-987-482-1
; Sequence 1, Application US/09987482
; Publication No. US20020184656A1
; GENERAL INFORMATION:
; APPLICANT: BHANDARI, POONAM
; APPLICANT: SHASHIDHARA, L.S.
; TITLE OF INVENTION: IN VIVO ASSAY SYSTEM FOR SCREENING AND VALIDATION OF
; TITLE OF INVENTION: DRUGS AND OTHER SUBSTANCES
; FILE REFERENCE: 056859-0134
; CURRENT APPLICATION NUMBER: US/09/987,482
; CURRENT FILING DATE: 2002-03-21
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2843
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-987-482-1

Query Match 3.1%; Score 299.5; DB 9; Length 2843;
Best Local Similarity 18.9%; Pred No. 6.2e-08;
Matches 398; Conservative 270; Mismatches 728; Indels 709; Gaps 93;
QY 69 TKRSLOESTRFSQL--VEBLKIIICAFOLD-----TGLE-----YAN 103
Db 874 SKRGLQISTTAAQIAKVMEEVSAIHTSQEDRSSGSTTLLHCVTDERNALRRSSAAHTSN 933
QY 104 SYNFAKKENNSPEHLKDEVSIIOSMGYRNPRAKRLLOSEPENSLOETSLSVOLSLGTVR 163
Db 934 TYNFTKSENSNR-----TCSMPY-----AKLEYKRSSNDLSNVSSSDGYGRGQM- 979
QY 164 TLRTKRIQIPQKTSVVVIELGSDSSDENVKATYCVSGVDQELLOITPQGRDEI-----216
Db 980 -----KPSI-----ESYSEDESK--FCSYGY-----PADLAHKIHANHMD 1015
QY 217 ----SLDSAKKAACFFSETDVTNTEHHQPSNNDLANTTEKRAAE---RHPEKYQGSVSNI 269
Db 1016 DNDGELDTPTNLSKYSDQOL-NSGRQSPSQNERWARPKHIIIDEIKQEQQRQSNQSTT 1074
QY 270 HVEPCGTNTHASLQHENSSLLLTDRMNVKAEFCNKSQKQGLARSHQNRWAGSKETCN 329
Db 1075 Y----PVYTESTDDKH-----LKFQPHFGQOECVSPYRSRGANGSETNR-VGSNHGIN 1122
QY 330 DRRTPTSTKKVLDLNDAP-----LCERKEWKNQKLPCE-----ENPRD-----366
Db 1123 QNVSQLCOEDDYEDDKPTNYSERYSEEEQHEEERPTNYSIKYNEEKRHVDQPIDYSLK 1182
QY 367 -TEDVPWITLNSIOKVNWFERSD-----ELGSDSDHGESESNKAVADVLDVLE 418
Db 1183 YATDIP-----SSQKQSPFSKSSSGQSSKTEHMSSESTSTPSSNAKQ-----NQ 1230
QY 419 VDEYSGSEKIDLLASDPHEALICKSERVHKSVE-----SNED 458
Db 1231 LHPSSAQSR-----SQQPKAATCKVSSINQETIOTYCVEDPTICFSRCSSLSUSSAED 1285
QY 459 KIFGTYRKAKSLPNLSHVTE-NLIIGAP-----VTEPQIIQERPLTNLKKRRKPTSGI 512
Db 1286 ELGCGNQTQEAADSANTLOIAEIKEKIGTSAEDPYSEVPVAVSQHPRT---KSRLOGSSL 1342
QY 513 HPEDFIKA-----DLAVOKTPEMINQGTNOTEQNGQVYMNITN 550
Db 1343 SSESARHKAVFEFSGAKSPKSGAQTPKSPPEHYVQETPLMFSRCTSVSSLD-----1394
QY 551 SGHENKTGDSIQNEK-----NPNPIESLEK-----SAFKTKAEPISISSINMELE 597
Db 1395 -SFESRSTASSVQSPFCSCGMVSGIISPSDLPSQGTMPPSRSKTPPPPPQTAOKREVP 1453
QY 598 LNIHNSKAPKNNLRKRSSTRIHALELWVSNLSPN-----CTELQID--SCSS-- 646
Db 1454 KN-----KAPTAE--KRESGPKOA-AVNAAVQVQLVDPDADTLLHFATETPDGFCSSSL 1506
QY 647 -----SEEKKKKYNO-----657

Db 1507 SALSLEPFIOKDELRIIMPVQENDNGNETESEQPKESNENQKEAKETIDSEKOLLDD 1566
 QY 658 -----MPVRISRNLOLMGEPATGAKK-----SNKP----- 684
 Db 1567 SDDDDIEILEECIISAMPTKSR-----KAKPAQATASKLPPPVARKPSOLPVYKLLPSQ 1621
 QY 685 NBQTSKRHDS-----DTPPELK-LTNAPGSFTKCSNTSELKEFVNPSLPREEKEKLETVK 739
 Db 1622 NRLQPKHVSTPGDDMPRVYCVETGPIINFSTATSLSDL----- 1660
 QY 740 VSNNAEDPKDMLSGERVLQTERSESSISLVPTGYTQES-----ISLLEVSTLGKAK 795
 Db 1661 ---TIESPPNELAAGEVGRGGAQSSEFEKRDITPEGRSTDEAOGGKTSVTPIDELDNK 1717
 QY 796 TEPNKCVCQAAPENPKLIH-----GCKDNRNDETEGFK-----Y 831
 Db 1718 ABECDILAEICINSAMPKCKSHKPFVRKKIMDQVOQASASSAPKNQLDGKKKPTSPVK 1777
 QY 832 PLUGHEVNH-SRETSTEMESELDAQYLQNTFKVSKROSFALFSN-----PGNAEEBCAT 884
 Db 1778 PIPQNTYRTRVRKNADSKNMLNAERFSDNKKQNLKNNKSKDFNDKLPNNEDRVGRS 1837
 QY 885 FSAHS-----GSL-----KQSPKVTFE 902
 Db 1838 FAFDSPHHYTPIEGTPYCFSRNDSLSLDDDDVDLSREKAELKAKENKESKAVTSH 1897
 QY 903 CEQKEENGKGNESNIKPVQTVNITAGFPV-----GOKDRP-----VD 940
 Db 1898 TELTSNQSANKTOAIAKQPIRQNPQPILOKQSTFPQSSKPIDRGAATDEKLQNAFIE 1957
 QY 941 NAK-CSIKGGSRFLCS-----SQFRGNETGLIT-----PNKHGLLQNPYRT--- 980
 Db 1958 NTPVCFSHNSLSLSDIDQENNNKENEPKETEPDPSQSGEPKQASGVAPKSFHVEDT 2017
 QY 981 PLFP-----IKSFVTKCKKNLLEENFEHEHSPEREMG-----NE 1017
 Db 2018 POCFSRNSLSLSDIDEDDLQBCISSAMPKKKPSRLKGDNKHS---PRNMGILGE 2074
 QY 1018 NIPSTVSTIISRNNTRENVFKREASSNINEVGSSTNEVGSINEIGSSDENIQAEIENRG 1077
 Db 2075 DTLDLKDIOIPDSEHGLSPDSEFMDKATQEGANSTVSLHQAAAA-----ACLSRQAS 2129
 QY 1078 PKLNAMRL--GV-----LOPVYKQSLPGSKHPEIKQYEYEVQVTVNTDF--- 1124
 Db 2130 SDSGILSLKSGISLSPFHUTPD--OEEKPTNSKNGPRILKPEKSTLTETKIESESKG 2187
 QY 1125 -----SPYLIDNLBQPMGSSHASOVCSFETPDLLDDGEIKEDTSAE 1167
 Db 2188 IKGKKVYKSLITCKVRSNSEISGOMQPIQANMPS-----ISRGRTMHIHPGV 2237
 QY 1168 NDIKESAVFSKSVQ-KGELSRSPSPTHLAQYRGAKKLESSEENLSSEDEELPCF 1226
 Db 2238 NSSSTSPVSKKGPPLTPASKSPS---EGQTATSPRGAPKSVKSLSPVAR----- 2287
 QY 1227 QHLLFGKVNIPQS--TRHST---VATECLSKNTEENLLSLKSLNDSNOVILAKASQE 1282
 Db 2288 QTSQIGGSKAPSGSRDSTSPSPAQOPLSRPIQS---PGRNISPGRNGI-----SPP 2339
 QY 1283 HHLSEETKCSASLFSOCSELEDITANTQDPFLIGSSK-----QMRHQSSEQGVG 1334
 Db 2340 NKLSQLPRTSS-----PSTASTKSS-----CSGKMSVTSPCGRQMSQNLTKQTG 2383
 QY 1335 LS-DKELVSDDEERGCTGLEENQ-----EQGMSDNLGEAASCESTSVSEDCSLG 1385
 Db 2384 LSKNASSIPRSESASKGLQNMNGNANGANKKVELSRMSST---KSSGESDRS---ERPVL 2437
 QY 1386 SSQSDILTOORDTMOHNLKLOQEMAEALVLEQHSQP-----SNSYPSIISDSALE 1440
 Db 2438 VRQSTFEKAPSPILRKL-----EESASFEL--SPSSRPASTRQOQTPVLSPSL--- 2488
 QY 1441 DLRNPEOSTSKAVLTQKSEYPISONP-----EGLSADKFEVSADSTSKNKEFCVER 1495

Db 2489 ----PDMSLTHSSVQAGGWKRLPPNLSPTIEYNGDRPAKRHRDIARSHSESPSRLP-INR 2543
 QY 1496 S-----SPSKCPSLDDRWYMHSCSGLQNR 1520
 Db 2544 SGTWKREHSHKSSSLPRVSTWRRRTGSSSSILSASSESEKAKSEDEK-HVNSISGTQKSK 2602
 QY 1521 -NYPQBELIKVVDVEE-----QOLEESQPHDLTETSYLPRODLEGTPLYESGISLFS 1572
 Db 2603 ENQVSAKGTWRKTKENEFSTNSTQTVSSGATNGAESKTLIYQMAPAVSKTEDVWVRIE 2662
 QY 1573 DDPESDPSEDAPESARVGNIP---SSTALKVPQLKVAESAQ----- 1612
 Db 2663 DCPINPRSGRSP---TGTPPVIDSVSEKANPNIKSDQAKQNVGNGSVPMRTVGL 2718
 QY 1613 -----SFAAHTTDTAGYN-----AMEESVSRKPELTASTERVNKRMSWV 1654
 Db 2719 ENRLNSFIQVADPQDKTEIKPGQNNPVVSETNESSIVERTPFSSSSSKHSPSGTVA 2778
 QY 1655 SGLTP 1659
 Db 2779 ARVTP 2783

RESULT 15

US-08-681-219-32
 ; Sequence 32, Application US/08681219
 ; Patent No. US20020058607A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Takaaki Sato and Junn Yanagisawa
 ; TITLE OF INVENTION: COMPOUNDS THAT INHIBIT THE INTERACTION BETWEEN
 ; SIGNAL-TRANSDUCING PROTEINS AND THE GLGF
 ; TITLE OF INVENTION: (PDZ/DHR) DOMAIN AND USES THEREOF
 ; NUMBER OF SEQUENCES: 35
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham LLP
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/681,219
 ; FILING DATE: 22-JUL-1996
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 0575/48962/JPW/JRM
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 278-0400
 ; TELEFAX: (212) 391-0525
 ; INFORMATION FOR SEQ ID NO: 32:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2843 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-08-681-219-32

Query Match 3.1%; Score 298.5; DB 8; Length 2843;
 Best Local Similarity 19.5%; Pred. No. 7e-08;
 Matches 362; Conservative 275; Mismatches 662; Indels 561; Gaps 98;
 QY 69 TRKSLOESTRFSOL--VEELLKIIICAFOLD-----TGLE-----YAN 103
 Db 874 SKRGLQISTAAQIAKMEVSAIHTSQEDRSSGSTTELHCVTDERNALRRSSAAHSHN 933

Qy	104	SYNFAKKENNSPEHLKDEVIISIQSGMYRNRAKRLLQOSEPMPISLOEYSLSVOLSNLGTVR	163
Db	934	TYNFTKSENRR-----TCSMPY-----AKLEYKRSSNDLSNVSSSSDGYGKRCQM-	979
Qy	164	TLRTKRIQIPQKTSVYIELGSDSSDTEYNKATYCSVGDOELLQITPOGTRDEI-----	216
Db	980	-----KPSI-----ESYSEDESK-----FCSYGQY-----PADLAHKHTHSANHMD	1015
Qy	217	-----SLDSAKKAACEFSETDVTNTEHHOPNSNDLNTTEKRAAE-----RHPEKYQCGSSVNL	269
Db	1016	DNDGELDTPTINYSLKYSDBQL-NSGRQSPSONERWARPKHIIIDEIKOSEQRQRNOSTT	1074
Qy	270	HVPEPCGTHHASSLOHENSLLLTKDRMNVEKAEFCNKSQOPGLARSOHNHRWAGSKETCN	329
Db	1075	Y-----PVYTESTDKH-----LKFQPHFGQECVSPYRSGANGSETNR-VGSNHGIN	1122
Qy	330	DRRTPSTEKKVDLNADP-----LCERKEWNKKOLPCS-----ENPRD-----	366
Db	1123	QNVSOLCQODDYEDDKPNYSERYSEEEQHEEERPTNYSIKYNEKRHVDDQIDYSIL	1182
Qy	367	-TEDVPWITLNSIQKVNWFERSD-----ELGSDSDHGESENAKVADVLDVNE	418
Db	1183	KATDIP-----SSQOKSFSEKSSGGOSSKTEHMSSESSENTSPSSNAKQ-----NQ	1230
Qy	419	VDEYSGSEKIDLLASDPHEALICKSERVHKSVSERN-IEDKIFGTYRKKASLPNLSHV	477
Db	1231	LHPSSAQSR-----SGOQKAATCVSSINOETIQTYCVEDTPI-CFSRCSLSLSLSSA	1283
Qy	478	TENLIICAFVTEPQIIQERPLNKLKRKRRPTSGLHPDEFIKKADLAVQKTPMINOGTN	537
Db	1284	EDE-IGC-----NOTTQEADSANTLO-----IAETKEKI--GTR	1314
Qy	538	QTEQNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPTSSSSISNMELE	597
Db	1315	SAED--PVSEVPVAVSHPTKSRLOQ-----SSLSESA-RHKAVEFSS-----	1356
Qy	598	LNTHNSKAPKNNLRR-KSSTRH-----THALEVYVRNLGPPNCPNTELOID	642
Db	1357	-----GAKSPSKGAQPTPKPPPEHYVQETPLMFSRCTSVSSLDSPESRISIA-----SSVOSE	1408
Qy	643	SCS-----SSEEIKKKYNOMPVHRSH-----NLQLMGEKAPATCAKKNKP	684
Db	1409	PCSGMVSGIISPDLDSPGQTPMPFSRSKTPPPPTQATKREVPKNKAP-TAEKRESGP	1467
Qy	685	N-----EOTSKRHSDTFPELTLNAPGSFTKCSNTSELKEFVNPSPREEKEKLE	736
Db	1468	KQAAVNAQVRQVLPDADTLHFATESTPDGFS-CSSLSALSLDPEFFQKOVELRIMP	1526
Qy	737	TVKVSNN-----AEDPKOLMLSGERVLOTERSVESSISLVPGYDTQGESISLLEYSTL	791
Db	1527	PVQENDNGNETESEQPKSENQOEK-EAEKTIIDSEKDLL-----DSDDDDDIIELEECII	1580
Qy	792	GKAKTEPNKCVSOCA--AFENPKGLIHCSK-----DNR-----NDTEGFKYPL	833
Db	1581	SAMPTKSSRRKAKPAQTASKLPPPVARKPSQLPVKLLPSQNRLOPKQHSVTFPGDDMPR	1640
Qy	834	-----GHEYNHSRETSIEMEESELDQYLQNTFKVSKRQSFALFSPGN--ABEECATFS	886
Db	1641	VYCVETGPIINFSTATSL-----SDLIE-----SPNELAGEGVRRG	1678
Qy	887	AHSGSLKKQSPKVTFECEQKENOGKNESNIK-----PVQTVNITAGFPVVGQ	934
Db	1679	AQSGEFEKRD-TIPTEGRSTDEAQGGKTSSTVIPLELDDNKAEEGDILAEACINSAMP-KGK	1736
Qy	935	KDPVDNACKSTKSGSRFCLSSQFRGNETGLITPNKHGGLQNPRIPLPPIKSF-----	989
Db	1737	SHKPFVRVK-----IMDVOQAQASSAPKNKOL--DGKHKKPTSPVKPIPONTE	1784
Qy	990	VKTCKCKNLLLEN--FEHSHMSPREMGEMNIPSTVSTISRRNIRENVFKEASSNINEV	1047
Db	1785	YRTRVRKNADSKNNLNAERVFSDNKSSKQNL-----KNNSKD--FNDKLPNNEDRV	1834
Qy	1048	-GS-----STNEVGSSSINEIGSSDENITQAEGLGNRQPKNLAMRLG	1087

Search completed: January 22, 2003, 16:52:35
Job time : 77.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 16:44:08 ; Search time 34.5 Seconds
(without alignments)
5191.259 Million cell updates/sec

Title: US-09-734-672-4
Perfect score: 9649
Sequence: 1 MOLSALRVEEVONVINAMOK.....LYQCQLDLYLPQIPHSHY 1863

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73.*

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9642	99.9	1863	1 A58881	breast/ovarian can
2	4882	50.6	1812	2 I49350	breast/ovarian can
3	373	3.9	2447	2 T16870	hypothetical prote
4	365.5	3.8	3329	2 T42205	breast cancer susc
5	361.5	3.7	3343	2 T42207	breast cancer susc
6	358.5	3.7	5327	2 T13564	microtubule-associ
7	357.5	3.7	3329	2 T30904	breast cancer tumo
8	355	3.7	2954	2 T14156	kinesin-related pr
9	342	3.5	2271	2 F90073	hypothetical prote
10	341.5	3.5	1837	2 T41023	probable nuclear p
11	340.5	3.5	1837	2 T41023	364K Golgi complex
12	340	3.5	2829	2 A42771	reticulocyte-bind
13	339.5	3.5	3488	2 T34418	hypothetical prote
14	339	3.5	1957	2 T38077	hypothetical coile
15	334.5	3.5	3328	2 T30835	breast cancer tumo
16	333	3.5	6713	2 B89921	hypothetical prote
17	332	3.4	3225	2 I52300	giantin - human
18	330.5	3.4	3147	2 T18674	hypothetical prote
19	328.5	3.4	5170	2 T15367	hypothetical prote
20	327	3.4	1875	2 S38173	myosin-like protei
21	327	3.4	3418	1 G02334	breast cancer tumo
22	323	3.3	3259	1 A58539	giantin - human
23	322	3.3	1658	2 S55101	hypothetical prote
24	321.5	3.3	4377	2 A55575	ankyrin 3, long sp
25	320	3.3	1738	2 T14867	interaptin - slime
26	319	3.3	65	2 G02999	BRCA1 - rhesus mac
27	319	3.3	1410	1 A57013	early endosome ant
28	318.5	3.3	1526	2 A45605	mature-parasite-in
29	317.5	3.3	3507	2 T34513	hypothetical prote

Query Match 99.9% Score 9642; DB 1; Length 1863;

30 317 3.3 2346 2 T13829
31 316.5 3.3 2484 2 T26216
32 316.5 3.3 2607 2 T26215
33 316 3.3 1727 2 T50073
34 315 3.3 2481 2 D90011
35 314 3.3 1871 2 D96796
36 313.5 3.2 1679 2 S48385
37 312.5 3.2 1790 2 S67593
38 312 3.2 2938 2 T30249
39 312 3.2 3924 2 S37431
40 311.5 3.2 2748 2 S57976
41 310.5 3.2 2845 2 I49505
42 309.5 3.2 2253 2 T30336
43 304.5 3.2 1744 2 JH0720
44 301 3.1 1269 2 F84730
45 301 3.1 2464 1 QRMSPI

ALIGNMENTS

RESULT 1

A58881
breast/ovarian cancer susceptibility protein BRCA1 - human
C:Species: Homo sapiens (man)
C:Date: 16-Oct-1998 #sequence_revision 16-Oct-1998 #text_change 18-Jun-1999
C:Accession: A58881; A54652; I58130
R:SKolnick, M.H.
submitted to GenBank, September 1994
A:Description: Human breast and ovarian cancer susceptibility (BRCA1) mRNA, complete
A:Reference number: A58881
A:Accession: A58881
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1863 <SKO>
A:Cross-references: GB:U14680; NID:955931; PIDN:AAA73985.1; PID:955932
R:Wiki, Y.; Swensen, J.; Shattuck-Eidens, D.; Futreal, P.A.; Harshman, K.; Tavtigian, ye, C.; Hattler, T.; Phelps, R.; Haugen-Strano, A.; Katcher, H.; Yakumo, K.; Gholami, ow, P.K.; Norris, F.H.; Helvering, L.; Morrison, P.; Rostock, P.; Lai, M.; Barrett, J Science 266, 66-71, 1994
A:Authors: Lewis, C.; Neuhausen, S.; Cannon-Albright, L.; Goldgar, D.; Wiseman, R.; K A:Title: A strong candidate for the breast and ovarian cancer susceptibility gene BRC A:Reference number: A54652; MUID:95025896; PMID:7545954
A:Accession: A54652
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-444, 'D', '446-487', 'S', '489-593, 595-600, 'M', '601-1452, 'VLQ', '1456-1471, 'X', '14 A:Cross-references: GB:U14680
R:Hosking, L.; Trowsdale, J.; Nicolai, H.; Solomon, E.; Foulkes, W.; Stamp, G.; Signe Nature Genet. 9, 343-344, 1995
A:Title: A somatic BRCA1 mutation in an ovarian tumour.
A:Reference number: I58130; MUID:95315980; PMID:7795636
A:Accession: I58130
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1812-1813, 'PGQRTMASMOLGRCVRLHW' <HOS>
A:Cross-references: GB:S78558; NID:g1000568; PIDN:AAB34725.1; PID:g1000569
A:Note: frame shift mutation from an ovarian tumor
C:Genetics:
A:Gene: GDB:BRCA1
A:Cross-references: GDB:136611; OMIM:113705
A:Map position: 17q21-17q21
C:Superfamily: breast/ovarian cancer susceptibility protein BRCA1; RING finger homolo
C:Keywords: mammary gland; nucleus; ovary; tumor suppressor; zinc finger
F:20-70/Domain: RING finger homology <RNG>
F:24-64/Region: zinc finger C3HC4 motif
F:607-621/Region: nuclear location signal
F:651-664/Region: nuclear location signal
F:679-692/Region: nuclear location signal
F:24,27,44,47/Binding site: zinc (Cys) #status predicted
F:39,41,61,64/Binding site: zinc (Cys, His, Cys) #status predicted

Db 1754 VSAPTKITAVPNVAAGPHQNOQQLYSDMAQAQSTPIRTTPQPGTGSAPQAQTPQSHLAQ 1813
QY 1590 VGNIPSTSAKVPOLK-----VAESAQSPAAAHTTDTAGYNAMESVSRE 1635
Db 1814 LGOFVNGANOQAQPOQGMYTAAQLQAMQAQAQAQAQAQAAAE--AAYOAQVAQAQARA 1872
QY 1636 KP 1637
Db 1873 AP 1874
RESULT 4
T42205
breast cancer susceptibility protein BRCA2 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T42205
R:McAllister, K.A.; Haugen-Strano, A.; Hagevik, S.; Collins, N.K.; Brownlee, H.; Futreal
submitted to the EMBL Data Library, February 1997
A:Description: Characterization of the mouse and rat homologs of the BRCA2 breast cancer
A:Reference number: 222073
A:Accession: T42205
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3329 <MCA>
A:Cross-references: EMBL:U89652; NID:g2443438; PID:g2443439; PIDN:AAB71377.1
A:Experimental source: strain CDI; 129Sv; ICR Swiss
C:Genetics:
A:Gene: BRCA2
C:Superfamily: breast cancer tumor suppressor BRCA2
Query Match 3.8%; Score 365.5; Db 2; Length 3329;
Best Local Similarity 19.4%; Pred. No. 5.1e-08;
Matches 402; Conservative 290; Mismatches 750; Indels 635; Gaps 93;
QY 72 SLQESTRFSOLVELKLIIC-----AFQDPTGLEAYANSYNAFAKENNS---PEHLKD 120
Db 466 SLEDSTAGKQWVSRTSQAACLSPSIRKSIFKREPLDLETGLTVFSDSMSTNSTFEEHEAS 525
QY 121 EVSIQSMGYRNRAKRLQLOSEPENPS---LQETSLSVOLSNLGTVTLTQRIQPKT 176
Db 526 ACGILITACQBRDSTCPSSVDGTGSMPTTLTDTSATVK--NAGLSTLKNKR----- 577
QY 177 SVYIELGSDSEDTVKATYCSVGDOBLQITPQGTREISLDSAKKAACEFSETDV--- 233
Db 578 -----KFIY-SVSDDASLOGKKILOTHROLELNL-SAQLAASAFVPLT 619
QY 234 -TYTEHHOPNNDLNTEKRAAERHPEKPYOGSSVSNLHVPECGTNTHASSLQHENSSL-- 290
Db 620 FTNVNSGIPDSSD-----KKRCLPNDE-----EPSLTNSFGTATSKAISVIHA 663
QY 291 LLTKDRMNYEKAFCNKSQKOPGLARQHNRPWAGSKETC-NDRTPST---EKKVDLNA-- 344
Db 664 LISQD-LNDKEAIVIEKPOPYTAREADFLLCPLPERTCENDQSPKVSNGKEKVLVSACL 722
QY 345 -----DPLCRKEWNKOKL-PCSENPRDTEVPW----- 372
Db 723 PSAVOLSSISFESQENPLGDHNGTSLKLTSPSKLPLSKADWVSREKMKMPKELQCESC 782
QY 373 ---ITLNSSTQKYNWFSRDELLGSDSDSHDGESENAKV-----ADYLDVINEYDEY 422
Db 783 KVNIELSKNILEVNEICILSE-----NSKTPGLLPGENIEVASSMKSQ 827
QY 423 SGSEKIDLLASDPHEALICKSR-----VHSKSVESNIEDKI-----FGKTYRKASLP 472
Db 828 FNQNAKI-----VIQKQDGSGPFTISEAVNMNSELFPDPSGNNFAQVNTNCKNP 877
QY 473 NLSHVTENLIIGAFTVEPQIIQERPL--TNKLRKRRTPSGLHPEDFIKKAADLAVOKTPE 530
Db 878 DLGSSVE-----LQEDLSHTQGPGLSKLSPRAVDEVDAAHAQAQVLTIKDSD 924
QY 531 MINOGTNOTEQOVNMNITNSGHENKT-KGDSIQNEKNPNPNIESLEKESAFKTKAEP ISS 589

Db 925 SLAVVHDYTEKSRNNIEHQHQTEGDKDFKSNSSLNMMKSDGNSDCSDKWEF---LDPVLN 981
QY 590 -----SISNMELELNIIHNSKAPK--NLRKRKSTR-----HIHALBELVSRNLSP 634
Db 982 HNFSGSRTASNKEIKULSEHNKVKRSMFFRDIIEQYPTRLACIDIVNTLPLANOQKILSEP 1041
QY 635 NCTELQIDSCSSSEIEKKKYNQMPVRHSRNLQLMCKEKPATGAKKSNKPNEQTSKRHDS 694
Db 1042 HIPDLK-----SVTVSTQSHNQSSVSH-----EDT 1067
QY 695 DTPPELKLTAPOGFTKCSNNTSELKEFPNPSLPREEKEKLETVK-----VSNNAEDP 747
Db 1068 DTAPQM-LSSKQDFHNNLTTSQAEITELSTLLEESGQFEFTQFRKPSHIAQNTSEVP 1126
QY 748 KDLMLSERVLQI-ERSVESSSISLVPCTDYGTQESISLLEVLSTLGLKAKTPEPKNCVSQA 806
Db 1127 GNMV-----VLSTASKEMKTDLHLVPDPSVGOTDHSKQFEGSA-GVKQFPFHLLDTCN 1181
QY 807 AFENPKGLIHGCKSDKNDNDE--GFKYPLGHEVNHRSRETSIEMEESLDAQYLQNTFKVS 864
Db 1182 --KNT-----SCLPNINEMEFEGFCSALGTKLSVSN-----A 1213
QY 865 KROSFALFSPGNABEECA-----TFSAHSGS-----LKKQSPKVTFE---CEQK 906
Db 1214 LRKAMKLFSDIENSEEPSAKVGRGFSSSAHDSVASVEFKIKQNTKESFDEKSKQCVT 1273
QY 907 BENQ-----GKN-ESNIKPVQ-----TVNITAGFPVV---GQK 935
Db 1274 LONNIEMTTCIFVGRNPEKIKNTKHEDSYTSSQORNLENSDGMSMSTSG-PVYIHKGDS 1332
QY 936 DKPVD-NAKCSIKGGRFCLSSOFRGNETGLITPNKHGLLQNPYRIPPLPIKSFVKTKC 994
Db 1333 DLADQSKC-----PESC--TOVARENTQIKENISDLTCLLEIMKAEETCMKSKDKQL 1385
QY 995 KKNLLEENFEHSMSPERMGNEIPSTVSTISRNIRENVFVREASSNINEVSGSTNEY 1054
Db 1386 PSDKMEONIKEFNTS-----FQTASGKNTR--VSKESLNKSNIENRETDE- 1429
QY 1055 GSSINEIGSSDENIQAEQLGRNRPKLNAMLRLGLVQPEVYKQSLPGSNCKHPETK-QOEY 1113
Db 1430 ---LTVISDS-----LNSKILHGINKDKMH-----TSCHKKAISIKKVF 1465
QY 1114 EEVYQTVNTDFSP-----YLISDNLEQPMGSSHAS-----QVCSETPD---DLDDGE- 1158
Db 1466 EDHPIVTVSOLPAQQAQHPYEIESTREPTLLSHTASGKVKIMQESLDKVNLFDETQY 1525
QY 1159 IKEDTSPAEND--IKESSAVFSKVQKGLSRSPSPPTHLAQGYRRGAKKLESSENL 1216
Db 1526 VRKTASFQSGSKPLKDSKELTLAYEKIEVT-----ASKCEEMQNFV 1567
QY 1217 SSEDEELPCFQHLLFGKVNNIQSTSRHSTVATE-----CLSKN---TEENL 1260
Db 1568 SKETEMPLQOQNYHMYROTENLKTSGTSSKVQENIENNVEKNPRICICOSSYPVTEDSA 1627
QY 1261 LS--LKSLNDCSNOVLAK-----TAN-----TNTQ-----DPFLIGSK-- 1322
Db 1628 LAYTEDSRKTCVRESSLSKRWLRQGGKLGTRNTIKIECVKEHTEDFAGNASEYHSL 1687
QY 1285 -----LSEETKCSALFSQCSLEDL----- 1306
Db 1688 VIIRTEIDTNHVSQVSTLLSDPNVCHSVLSQSSFCCHDMHNDSGYFLKNKIDSDVPP 1747
QY 1307 -----TAN-----TNTQ-----DPFLIGSK-- 1322
Db 1748 DMRNAEGNTISPRVSATKERNLHPQITINEYCVQKLETNTSPHANKDVAIDPFLDLSRCK 1807
QY 1323 -----QMRHQSQGVGLSDKELVSDDEERGTCLEBNNOEEOESMDSNLGEAASGESET 1376
Db 1808 VGSIVFTIASHQETE----RTKEIVTDNCYKIV-----EONRQSKPDTCOTSCHKVL 1855
QY 1377 SVSEDCSGLSSQSDILTTOQORDTM-QHNLIKLOQEMAEVLEQHGQSOPSNYPSTISD 1435

Db 1856 DSKDFICPSSGDVGINSRKDSFCPHNEQILQHNS-----MFLGKKAATPPVGLT 1908
Qy 1436 SSALEDLRNPEQ-----STSEKAVLTISOKSSSEYPISONPEGLSADKEEVSADS 1483
Db 1909 WDTSKSIRPEPPQAHPSTRYGYGFSTAGSKAIQVSDASLE-KARQVSEMDGDAKQLSSWV 1967
Qy 1484 STSKNKEP--GVERSSPSKCPSLDDRWYHSCSGSLQ-----NRNYP-----1523
Db 1968 SLEGNEKPHHSVKRENS-----VVHSTQGVLSPKPLPGNVNSSFVSGFSTAGGK 2017
Qy 1524 ----SOBELIKVVDV-----EBQOLEESG-PHDLTFTSYLPQDLBGTYP--LESGIS 1569
Db 2018 LVTVSEALHVKMGLEEDFLIRTEHTLQHPIDEDVSKILPQPCAEIRTPYVNSKLQ 2077
Qy 1570 LFSDDPESDPSDRAPESARVGNIPSTGALKVPOLKVAESAQSPAAHTTDTAG-----1624
Db 2078 KTYNDKSSLPSNVK--ESSSGNTQSIIEVSLQISQWERNQDTQLVLGTVKSHSKANLLGK 2135
Qy 1625 YNAMESVSREKPEL-TASTERVYNKRMVSVGLTPEEFMLVYKFARKHHITLTNLITEE 1683
Db 2136 EQTLPNQIKVKTDEMKTFSDPVVKTNVGEYSKESENYFETEAVESAKAFMEDDELTDSE 2195
Qy 1684 TTHVWMTDAEFVCERTLYFIYGIAGCKVWVYFWVTQSIKERKMLNEHD--FEVRGDVV 1741
Db 2196 QTHAKCSL---FTCPQNETLFNSRTRKRGGVTVDVAGOPPDKRSLNEDFDRIIESKGRSL 2252
Qy 1742 NGRNHQGPRARES---QDRKIF--RGLEICCYGPF 1772
Db 2253 T-----PKSTPDGTGVKDRSLTFTHMSLEPVTGPF 2283

RESULT 5
T42207
breast cancer susceptibility protein BRCA2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T42207
R:McAllister, K.A.; Haugen-Strano, A.; Hagevik, S.; Collins, N.K.; Brownlee, H.; Futreal
submitted to the EMBL Data Library, February 1997
A:Description: Characterization of the mouse and rat homologs of the BRCA2 breast cancer
A:Reference number: 222073
A:Accession: T42207
A:Status: preliminary; translated from GE/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3343 <MCA>
A:Cross-references: EMBL:U89653; NID:G2443440; PID:G2443441; PIDN:AB71378.1
A:Experimental source: strain Sprague-Dawley; testes
C:Genetics:
A:Gene: BRCA2
C:Superfamily: breast cancer tumor suppressor BRCA2

Query Match 3.7%; Score 361.5; DB 2; Length 3343;
Best Local Similarity 19.2%; Pred.No. 7.7e-08;
Matches 379; Conservative 297; Mismatches 715; Indels 579; Gaps 96;

Qy 66 NDTIKRSLOESTRFSOLVBELKIIICAFOLDTGLEYANSYNAFFAKKENNSPEHL-KDEYSI 124
Db 1087 HNLTPSQAEITELSTILE-----ESGQFEFTQPKN--PSHIAQNTSA 1129
Qy 125 IOSMGYRNRAKLLOSEPNNPSLQTSLSVLSLGTVTRTLTKRIQIOPKTSYVIELGS 184
Db 1130 V--LGNMAVVRTASEEWKVDLH---LPLNPSSVGQIDHNKKFCELVGVKQS-----SS 1179
Qy 185 DSSEDVTKATCSVGDQELL---QITQGTDEISLSAKKAACEFSETDVTNTE-----237
Db 1180 HLLDTCNNTSCFLPIKEMFEGGFCGALGKLSVSNALRAMKLFSDIENISEEPSTK 1239
Qy 238 -----HH-----QPSNNDLNTTEKRA-----AERHPKY 261
Db 1240 VGRGFSSCAHDSVASVFKIKKQNTKDSFDEKSSKQVTVQNNKEMTTCILVDENPENY 1299
Qy 262 QGSSVSNLHVEPCGNTWASSLQHSNLSLLTKDRMYEKAFCNK--SKQGLARSOHN 319

Db 1300 ----VKNIQDNNTYTCSORNAYKLENSDV-----SKSSTSGTVYINKGSDLP-FAAEKGN 1350
Qy 320 RWAGSKETC---NDRRTPTSEKKVDL-----NADPLCERKEWNKKQLPCSPENRPTD 369
Db 1351 KY---PESTQTVREENAQIKESVSDLTCLVYMKAEETCHMKSSDKEQLPSDKMEQNMK 1407
Qy 370 VPWITLNSIQKVNWFSRSDDELLGSDSHDGESNAKVAVDLVNLNEVDEYSGSSEKI 429
Db 1408 ----FNLSFQ-----TASGNIRVSKESLNKSVNILD--QETEDLTVTSDSL 1448
Qy 430 DLLASDPHEALICKSER-----VHKSVSNEIDKIFOKTY--RKKASLPNLNLSHVTE 481
Db 1449 ----NSKILCGINKDKMHISCHKKSI--NIK-KVFEHFPPIGVTSOLPALQOQPEYE 1498
Qy 482 IIGAFVTPP-----QIIQE-----RPLTNKLKRKRRTPSGLHP-----E 515
Db 1499 I--ESIKEPTLLSFHTASGKKVKVIMQESLDKVKNLFDQTYVRKTTNFGHESKPKLDRE 1556
Qy 516 DF-----IKKDLAVQKTPPEMINQNTQE-----QNG-----QVMNITNSGHENKTKGD 560
Db 1557 DYKERLTLAYEKIEVTASKCEMQNFVSKQTEMLPQONDHMYROTENLTSNGSSPKVHG- 1615
Qy 561 SIQN--ERNPNPIESLEKSAFKTKAEPIS-----SSISNMELELNINHSKA 605
Db 1616 NIENKIEKNPR-ICCIQSSVFTEDSALACYTGDGRKTCVGESSLKGGKWLREQSDKL 1674
Qy 606 PKKNLRKRKSTRH-----IHALELVVSRNLSPPNCTELQIDSCSSSEELKKKYNQ 657
Db 1675 GTRNTIETQCKVEHTEDFAGNALYEHSLVIIRT-----EIDTSHVSENOAQLFSD 1725
Qy 658 MPVRHRLQLMEGKEPATGAKKSNKPNEQTSKRHSDSTFPPELKLNTNAPGFTFKCSNTSE 717
Db 1726 PNVCHS----YLSSHSCFCHDDMHNDSGYFLKDKIDSDVQPMKNTGNAIPFKISATKE 1781
Qy 718 LKEFVNPSLPREEKEKLETVKVSNNAEDPKDLMLSGERVLTQTERSVESSSISLVPCTDY 777
Db 1782 IK--LHPQTVNBEQVKLET-----NASPYANKNTAIDSAMLDLRCKNVGSPVFI--TTH 1832
Qy 778 GTOESISLLEVTGLGKAK-----TEPNKCVSQC-AAFENPKGLIHGCSKON--RNDTE 827
Db 1833 -SOETVRMKEIFTNCNKIVONERESKPDTCQTSCHKALDNSEDFICPSSSGDVCINSPM 1891
Qy 828 GFKYPGLGHEV-NHSRETS-----IEMESELDAQYLQNTFKVKSRQSFALFS---873
Db 1892 AIFYPQSQILQHNQSVSLGKKAATPPVSLTWTDKTSIRGSPQEVHPRTYGFSTASC 1951
Qy 874 ----NPGNAEECATFSAHSGSLKKQSPKVTFECEQKEENOGKNESNI-----K 918
Db 1952 KAVQVSDASLEKARQVSEIDGDAKQLASWVSLGNEKSHHSVKRESSVYVHNTHGVLSLR 2011
Qy 919 PVQTVNIT---AGFPVVGQDKPV-DNAKCSIKGSRFCLSSOPFRGNETGLITPNKHGL 973
Db 2012 KTLPCNVSSFVSGFSTAGGKLVTVSESALHKVKG-----MLEEFD-----LIRTEHL 2060
Qy 974 LQNP-----YRIPPLPIKSFVKTKCKKNLLFENFEHSHMSPEREMGNENIPSTVSTIR 1028
Db 2061 QHSPTPDVSKIPP-----QPCLESRTPEYSVS-----SKLQTYNDKSR 2100
Qy 1029 NNIRENVFEKASSNINEVGSSTN-----EVGSSINIEGSSDENIAELGRNRPKLNAML 1084
Db 2101 -----SPNSYKESGSGNTQSLVSPQLSQMERKQET-QSVLGTKVYSQRKTNIL 2148
Qy 1085 RLGVLPQEVYKQSLPGSNCKHPEIKKQVEEVVQTVNTDFSYLLISDNLEQPMGSSHASQ 1144
Db 2149 E-----KKONLP-QNIKIESNKMETFSDVSMKTN-----VGE 2179
Qy 1145 VCSETPDLLDDGETKEDTSPAEND-IKESSAVFS-----SVQKGLSRSPSPFTHTL 1198
Db 2180 YYSKEPENYFETEAVEIAKAFMEDDELTDSEQTHAKCSLFCAPQNEALLNS-----RTRK 2234
Qy 1199 AQGYARGAKKLESSEENLSSEDEELPCQHLLFGKVNNI-PSQSTRHSTVATECLSKNTE 1257
Db 2235 RGMAGVAVGQPPIKRSLNE-----FDRIIESKGLSLTPSKSTPDGTIKORRLF----2284

Db 1214 LRKAMKLFSDIENSEEPAKVGPRGSSAHSDSVAFKIKKQNTKSTDEKSSKQCVT 1273
QY 907 EENO-----GKN-ESNIKPVQ-----TVNITAGFPVV--GQK 935
Db 1274 LQNNIEMTTCIFVGRNPEKIKYKTHEDSYSSORNNLENSDGSMSSTSG-PVYIHKGDS 1332
QY 936 DKPVD-NAKSISGGRFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLPIKSFVKTK 994
Db 1333 DLPADGSKC-----PESC--TOVARENTQIKENISDLTCLETMKAEETCMKSSDKQL 1385
QY 995 KKNLEENFEHSMSPERGMENISTVSTISRNNIRENVFKREASSNNINEVGSSTNEV 1054
Db 1386 PSDKMEONIKFEINIS-----FOTASGKNTR--VSKESLNKSNIFNRFDE- 1429
QY 1055 GSSINETGSDENITQABELGRNRPKLNALRLGVLOPEVYKQSLPGSNCKHPETK-KOEY 1113
Db 1430 --LTVISDS-----LNSKILHGINKDKMH-----TSCHKKAISIKKVF 1465
QY 1114 EEVVQVNTDFSP-----YLISDNLEQPMGSSHAS-----QVCSETPD---DLLDDGE- 1158
Db 1466 EDHFPVIVTSQAPQAQHPYEIESTKEPTLLSFHTASGKKVKIMQESLDKVKNLFDETOY 1525
QY 1159 IKEDTFAEND--IKESSAVFSKSVQKGLSRSPSPFTHLAQGYRGAKKLESSEENL 1216
Db 1526 VRKTASFSGSKPLKDSKKELTAYEKIEVT-----ASKCEEMQNFV 1367
QY 1217 SDEDEELPCFQHLIFGKVNINIPQSOTRSTVATE-----CLSKN-----TEENL 1260
Db 1568 SKETEMPLQOYHYMYRQTEENLKTNSGTSSKQVENIENNVKFNPRICCIQOSSYPVTEDSA 1627
QY 1261 LS--LKNSLNDSCNQVILAK-----ASQEH- 1284
Db 1628 LAYITEDSRKTCVRESLSRGRKWLREBQDKLGTNRNTIKIECVKEHTDFAGNASVEHSL 1687
QY 1285 -----LSBETCSASLFSQSCSEDL----- 1306
Db 1688 VIIRTEIDTHVSENQVSTLSDPNVCHLSQSSCFCHDDMHNDGYSFLKKNKIDSDVPP 1747
QY 1307 -----TAN-----TNTQ-----DPFLIGSK-- 1322
Db 1748 DMKNAEGNTISPRVATKERNLHPQITNEYCVQKLENTSPHANKOVAIDPSLLDSRNCK 1807
QY 1323 -----QMRHQSSQGVGLSDKELVSDDEERGTLGLENNOEBSMNSLNGEASGGESET 1376
Db 1808 VGSILVFITTAHSQETE-----RTKEIVTDNCKIV-----EQNRQSKPDTCOTSCHKVL 1855
QY 1377 SVSEDCSGLSQSDILTTQORDI-----MOHNLKLOQEALEAVLEHQHGSQPSNS 1428
Db 1856 DSKDFICPSSGDVCINSRKRKDFCPHNEQILQHN-----QSMS-----GLKKAAT 1901
QY 1429 YPSTIISDSSALEDLRNPQ-----STSEKAVLTSQKSSEYPISONPEGLSADK 1476
Db 1902 PPVGLTWDTSKSIREFPQAAPHSRYGIFSTASGKAIQVSDASLE-KARQVFSEMDGDA 1960
QY 1477 FEVSADSTSKNKEP--GVERSSPKCPSLDDRMVHSGSGSLQ-----NRNYP----- 1523
Db 1961 KOLSSMVSLGNEKPHHSVKRENS-----VHSTQGVLSLKPPLPGNVSSVFSG 2010
QY 1524 -----SQBELIKVDV-----EEQQLPESG-PHDLTETSYLPRLDLEGTPT 1563
Db 2011 FSTAGKLVTVSALHVKGMLEEDFLIRTEHTLGHSPIDEDVSKILPQPCAEIRTPY 2070
QY 1564 -LESGLISLFDSPEDSPEDRAPESARVGNIPSSGALKVPOLKVAESAQSPAAHTTDT 1622
Db 2071 PVNSKLQTYNDKSLSPSNYK--ESSSGNTQSIQVLSQSMERNQDQTLVLGTVKSHS 2128
QY 1623 AG-----YNAMESVSREKPEL-TASTERVKNRMSVWVSGLTPEEFMLVYKFARKHHITL 1676
Db 2129 KANLLGKEQTLQNIQVKTDEMKTFSQVVKTNVGEVYSKESSENYFETEAVERAKAFMED 2188
QY 1677 TNLITETTHVMKTDAEFVCEFTLYFLGIAGKQVWVFWTQSIKERKMLNEHD--F 1734
Db 2189 DELTDEQTHAKCSL---FTCPQNETLFNSRTRKRGGVTVDVAGQPPIKRSLNNEFDRII 2245

QY 1735 EVRGDVVNGRNHQPKRARES---QDRKIF---RGLEICCYGPF 1772
Db 2246 ESKGKSILT-----PSKSTPDGTVKORSLEFTHHMSLEPVTCTGPF 2283

RESULT 8

T14156

Kinesin-related protein - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000

C:Accession: T14156

R:Wood, K.W.; Sakowicz, R.; Goldstein, L.S.; Cleveland, D.W.

Cell 91, 357-366, 1997

A:Title: CENP-E is a plus end-directed kinetochore motor required for metaphase chrom

A:Reference number: Z17893; MUID:98028574; PMID:9363944

A:Accession: T14156

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2954 <WOO>

A:Cross-references: EMBL:AF027728; NID:g2586070; PID:g2586071; PIDN:AAC60300.1

C:Genetics:

A:Gene: XCENP-E

C:Superfamily: centromere protein E; kinesin motor domain homology

Query Match 3.7%; Score 355; DB 2; Length 2954;

Best Local Similarity 19.3%; Pred. No. 1.3e-07;

Matches 358; Conservative 276; Mismatches 678; Indels 546; Gaps 87;

QY 65 KNDITKRSLOEQRFSQVLEELKIIICAFOLDTG-LEYANSYNPAKKE--NNSPEH---- 117

Db 552 EKEIT--SLQQLQSKBE--BEKKELVQSFFELKIAELQESVAKNLEMTNRSIHNA 607

QY 118 -----LKDEVSTLIQSMGYNRNRAKRLQSEPNPSLQETSLS----- 153

Db 608 EVQTDVEKEVVRKEMSVGLDSGY-----NASNSDLQSDSVGDKRLUSSSHDECIH 657

QY 154 -----VOLSNLGVTRTLRTKRIQPKQTSVYIELGSDSSEDTVKNATYCVSGQDELL 205

Db 658 RKMLEQKIVDLEEF--IENLNKSKENDKQKS-----SEQDFMESIQLC-----EAI 701

QY 206 QITPQGTTRDISL--DSAKKAACEFSET---DVTNTEHHQPSNNDLNTTEKRAAERPEK 260

Db 702 MAFKANALELALMRONFONIIIE--NETLKREIADLERSLKNETNEFEILEKETQKE- 759

QY 261 YQGSYSVNLHVEPCGINTHASSLOHENSLLLTKDRMNVKEAFCNKSQPG-----LA 314

Db 760 -----HEAQLIHEIGSL-----KKLVENAEYMNONLEEDLETTKLL 796

QY 315 RSQHNWAGSKETCNDRRTPSTEKVV--DL-----NADPLCERKEWNKOKLPCSEN-PR 365

Db 797 KEQEIOLAEIR-----KRADNLQKKVNFPLSVSMGDSSEKLCEEIFOLKQSLSDAEAVTR 851

QY 366 DTED-----VPWITLNSSTQKVNWFSDLLGSDSDHSGSESNAK-----VADVL 413

Db 852 DAQKECSFLRSNELKEKEMEDTSNNYNOKEAASLFEKOLETEKSNYKMEADLQKELQ 911

QY 414 DVLNEVDYSGSSEKIDLLASDPHEALICKSE-----RVHSKSVESNIEDKIFGKYRKK 468

Db 912 SAFNEINYLNG-----LLAGKVPRLLSRVELEKVKYSEFSKQLEKALEEK-----NALENE 962

QY 469 AS-----LPN-----LSHVTENLIIGAFVTEPQIIQERPLTNKLRKRRTPTSGHL 513

Db 963 VTCLESEYKFLPNEVECLKNOISKASEIML-----LQKE-----GEH 999

QY 514 PEDFIKKADLAVOKTPEMINTQGTNQ--TEQNGQVMTINTSGHENKTKGDSIQNEKNPNPIE 572

Db 1000 SASIISKQEIIMQEQSEIQLTDEVTHTQSQVQOQTEEQYLEMKMKHDDL----- 1049

QY 573 SLEKESAFKTKAPISSSISNME-----LELNHNSKAPKKNRLRRKSSRTHIHAELVV 627

Db 1050 -FEKYTRNKSEADLLREMENLKGTMESVEVKIADTKHLELEETIRQKQLLH----- 1100

QY 628 SRNLSPNCTELQIDSCSSSEIEKKKKYNQMPVHRHNRNLQMLEGKEPATGAKKSNKPNQ 687
 Db 1101 -----EKKY-----FFQAMQTIPTPL----- 1118
 QY 688 TSKRHSDTFPELKLNAFGSTFKCNTSELKEFVNPSLPREEKEKLETVKVSNNADP 747
 Db 1119 -----SDSLPPSKLVG-----NSQDPTIENDYHNLIALATERNIMVCLTERNSLKE 1167
 QY 748 KDLMLSGERVLOTERSVESSISLPGTGYCTQESISLLEVLTLGKAKTEPNKVCQCAA 807
 Db 1168 QVIDLNTQLOAQSIKSDIQ-KPKQDEGEVKLLLEME----- 1208
 QY 808 FENPKGLIHGCSKDNDRNDEFGKPLGH-EVNHRET-SIEMESELQAOYLQNTFKVSK 865
 Db 1209 -----LLRGHLTDSQLSIE-KLQLENLEVTEKLTQLOEEMKNITIERNELQTNFEDLK 1260
 QY 866 RQSFALFNPNGNAEBECATFSAHSGSLKQSPKVFTECEQKEENGKGNESNTPVOTVNI 925
 Db 1261 -----AEHDSLKOD-----LSENIEQSIETQDE-----L 1284
 QY 926 TAGFPVVGOKDPVDNAK-----CSIKGGRFLCSQFRGNETGLITPNKHGLLQNPYRI 980
 Db 1285 RAAQELRQKQVDSFRQQLDCSV-----GISSPN-HDAVANQEKV 1326
 QY 981 PPLPIKSF-----VTKCKKNLLE-ENFEESHMSPEREMGNENIPSTVSTISR 1028
 Db 1327 -SIGEVNSLQSEMLRGERDELQTSKALVSELELLRAHVSVGE--NLEITKKLNGLEK 1383
 QY 1029 NNTRENVFKEASSNINEVGSSTNEVGSINIGSSDENIOAELGRNGRGPXNLAMLRGV 1088
 Db 1384 EILGSESEVLKMLENLKEDNNKLEQAEYSKQNFSLVEVFSQKL-----V 1436
 QY 1089 LOPEVYKSLPGSNCKHPKIKQYEEVQTVNTDFSYLLSDNLQPMGSSHASQVCE 1148
 Db 1437 DEIEVLKAQLKAAE-ERLEIKDQFELVQNTATN-----LVGKLETPLOADHE-----ED 1487
 QY 1149 TPDDLDDGEIK-----EDTSFAENDIKESSAFVSKVQKGLSRSPFTHLQAGY 1202
 Db 1488 SIDRRSEEMEIKVLGKERNQYLLERQEBKLELSNLEILOKEMETSVLLKDDLOQ-- 1545
 QY 1203 RGAKKLES-----SEENLSSDEDELPCFOHLLFGKVNINPQS 1241
 Db 1546 -----KLESLSSENILKENIDITLKHHSQDQALQKQQLQELQAKLNAIAASDNC--I 1598
 QY 1242 TRHSTVATCLSKNTPEENLLSKNSLDCSN--OVILAKASQEHLSSEETKCSA----- 1294
 Db 1599 TQKETISADCVIP-LEEKILLTEELHOKTNEQKLEHKELEQAQVELCEVHLKMS 1657
 QY 1295 LFSQCSSELDANTANTQDPFLIGSSQMR-----HOSESQGV 1333
 Db 1658 MTESK-SLESLOHEKHDTQQLLAKQMQVVVTOEKELQOQTHEHLTAEDVHLKNTIEL 1716
 QY 1334 GLSDRELVSDDERTG-----LEENNQEBS-----MDSNLGAAGSCSETSVS 1379
 Db 1717 GLNFK---NEAQOQTKKEQCLLENKELEOSQRLQCEIEELMKSLKDKESALETLKSE 1773
 QY 1380 EDCGLSSQSDILTTQ-----ORDTMOHNL-----IKLQQEMAELEAVLE 1419
 Db 1774 QKVINLQNMENMVMLEMLKNSQRTVTAERDQQLDDRESVEMSIETQDRLRAQEAQ 1833
 QY 1420 QHGS--QPSNYSPIISDSSAL-----EDLRNPEQST-SEKAVILT- 1456
 Db 1834 QOKDKVQELTSQISVLQEKISLENNQMLYNNVATYKTELSERDDLNQSKHLFSEITLISL 1893
 QY 1457 SOKSEYPIQNPGLSADKFPVSADSTSKNKEPGVRSRSPKCPSLDDRKY-----MHS 1512
 Db 1894 SLKEFEALEQ---AEKDKADAARKTIDITEKISNIEBQQLQOATNLKETLYERESLIQ 1949
 QY 1513 CSGSLQNRNYPQSEELIKVQVVE---EQOLEESGPH--DLTE-TSYLPQDLECTPYLE 1565
 Db 1950 CKEQLA-LNTEHLRETLKSKDIALKMEQDEERANKVIALTEKMSSLEEQINENVTTLK 2008
 QY 1566 SG-----ISLFSDDPESDPSEDRAPESARVGNIPSTSAKVPQLKVAESAQSPAAHTTD 1621

Db 2009 EGEGETFYLRPSKQSSSQSMEE-----LRESLTKDLQL-EBAKEISEATNE 2058
 QY 1622 -----TAGNNAMEESVSRKEPELT-ASTERVNRMS--WVVSGLTPEEFMLVYKFAKH 1672
 Db 2059 IKNLTAKISSLEETLQNASILNEAVSERENLRHKSQULVSEL--BQLSLTLK-SRDH 2113
 RESULT 9
 F90073
 Hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)
 C:Species: Staphylococcus aureus
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
 C:Accession: F90073
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O-
 ma, A.; Mizutani-Oi, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kato, C.; Sekimizu, K.
 Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
 Lancet 357, 1225-1240, 2001
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
 A:Reference number: A89758; MUID:21311952; PMID:11418146
 A:Accession: F90073
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2271 <KUR>
 A:Cross-references: GB:BA000018; PID:gl3702612; PIDN:BA043752.1; GSPDB:GN00149
 A:Experimental source: strain N315
 C:Genetics:
 A:Gene: SA2447
 Query Match 3.5%; Score 342; DB 2; Length 2271;
 Best Local Similarity 18.5%; Pred. No. 3.3e-07;
 Matches 310; Conservative 266; Mismatches 835; Indels 266; Gaps 49;
 QY 161 TVRLRTKQRTQPKTSVYIELGSDSEDTVVKATYCSVGDOELLOITPOGTRDEISLDS 220
 Db 638 TVTVVSTQANNKSTTTFTINV-VDTTAPTVT-----PIGDQSSSEVYSP-----ISP 683
 QY 221 AKKAACESETDVTNTEHHQPS-----NNDLNTETRAAERHPEKYGSSVSNLHVPE 273
 Db 684 IKIATQDNGNAVNTVTVGLPSGLTFDSTNTTISTG-----PTNIGTTSISIVSTDA 735
 QY 274 CGTNTHASLQHESSLLLTKDRMVERKAEFCNKQKGLARSOHNRWAGSKETCND--- 330
 Db 736 SGNKT-TTTFKYE-----VTRNSMSDSVSTSGTSQSSQSVSTSKADQSASSTSGSIV 789
 QY 331 RRTSTTEKVDLN-ADPLCEKERNKQKLPCESENPRDTPVWITLNSIQ-KVNEWFSR 388
 Db 790 STSASTSKSTVSLSDSVSASKSLSTSESNSVSSSTSTSLVNSQSVSSMSGSVSKSTSL 849
 QY 389 SDELLGSDSDHGESESNAKVADVLVDLNEVDE-----YSGSSEKIDLLASDPHEALICK 443
 Db 850 SDSISNSNSTEKSSELSST-SLSDSLRTSLSDSLSMSTSGSLSKSLSSTISIS----- 902
 QY 444 SERVHKSVEENIEDKIFKTYRKASLPNLSHVNTENLIIGAFVTEPQIIQPRPLTNLKL 503
 Db 903 ----GSSTASLSDST-SNAISTSTLSSESASTSDSISISANSQASQAS-----TSKSD 953
 QY 504 RKRPTSLGHPEDFIKKADLAVOKTPENINQGTNOEQNGQVMNITNSGHEKNTKGDSTQ 563
 Db 954 SQSTSIIS-----LSTSDSKSMSTSESLDSTSTSGVSGSLSTAAASQSVSTSDSMS 1006
 QY 564 -NEKNPNPIESLEKESAFKTAEPISSSISNMELELNHNSKAPKKNLRRKSRTHIHA 622
 Db 1007 TSEIVSDSISSTSGLSASDSKSMVSSSMST-----SQSGSTSESLDSQSTSDSDS 1058
 QY 623 LELVVRNLSPNCTELQI-----DSCSSSEIEKKKKYNQMPVHRHNRNLQMLEGKEPA 675
 Db 1059 KSLSLSTSQSGST 1118
 QY 676 TGAKKSNKPNQOTSKRHSDTFPELKLNAFGSTFKCNTSELKEFVNPSLPREEKEKL 735
 Db 1119 TASSESISQASASTSTSGSVSTSTSLSTNSERTSTSTSDSTSLSTSESDSI--SESTSTS 1176


```
Db 775 EQSNHIVNLIKSNLLLTYSVKOKLKADLNSLKSLSLQQDNFHMKAQIESNQBYTATV 834
QY 662 HSRNLQIMEGKEPATGAKKSNPNEQTSKRHD-SDFPELKLNTNAPGSP-----T 710
Db 835 DSMNSRILE-----LSNDRVANSKLSCSDVRLTLQN---SFDLREHQTILVLO 882
QY 711 KCSNTSELKEFVNPSLPR-----EKEEKLTVKVSNNAPDKDL 750
Db 883 LQSNITELKQDI--TLQRTVRNQLEIQTTELKERLKFMEERQENLQSKLIANA---NDT 936
QY 751 MLSGERVLQTSRVSSSISLVPQTDYGTQESISLLEVSTIGKAKTEPKVCVQCAAFEN 810
Db 937 TQNPDDNV-----EVEAISTEL-----ERTKEKLRAE---LEKSNIOQKYLASE----- 977
QY 811 PKGLIHGCKDNDRD-EGFKYPLGHEVNHRS--TSIE-----MEESLDA 854
Db 978 -KTL-----EMNTHQFQKHLVSEISTREKITSLSRSELLDNKRVEVLKEKESS 1030
QY 855 QYLQNTFKVSKROSFAFSGNPNABECCATFSAGSKLKQSPK---VTFCEQKEENQ- 910
Db 1031 KELAQLEDAVREKDSALSFKDKYKIRSDADRVITSUKEDIKERSLMKECHSNYSEI 1090
QY 911 ---GKNESNIKPVOT-----VNITAGFPV-----GOKD-----KPDVNAKCSIK 947
Db 1091 VSHGRTTQKLRDLRTFDEVNTKYLKLANFQOHSGLSGAEKDWNIOKAMEDISILK 1150
QY 948 G-----GSRFCLSSQF-----RGNETGLITPNKHGILLQNPYRIPPLPIKSF 989
Db 1151 DYTILGLENQKLLHSQFDSLQOITVLOQNSSENINISANLEAVOONDLR-----ELVSY 1205
QY 990 VTKCKKNLEENFEHSPEREMGNENIPVSTISRNINRENVEKASSN----- 1043
Db 1206 LRH-EKEIMDNKY-ELTILDNRGL-NOQVKSLOSTVDSLOLELNRLOSLPVSNDQTDTP 1261
QY 1044 -----INVGSTNEVSGSINIGSSDENIQIQLGRN 1075
Db 1262 IISGSOEQVLLYESNSVLKNDNAKLGKIQELEKEVEKELNASLNPLOTEINELKAEIGAK 1321
QY 1076 RG-----PKNLMLRLGLQVPEYVKOSLPSGNCKHPPEIKOEVEEVQ 1118
Db 1322 TASLNLKMEYNSRWKLRFOSVLNKNYERVDPTQLELKK-----NCEALEKEQOELTKIQ 1376
QY 1119 TVNTDFSYLISDNLEQPMGSSHASQVCSPTDDLLDQGEIKEDTSAENDIKESAVFS 1178
Db 1377 ETAKE-----TDTFKQOVNS-----LNEEVENL-----KKEVEQANTKNTRLAAAWN 1418
QY 1179 ---KSVQKGLSRSPFTHLQGYRGAKKLES---SEENLSSEDEELPCQOHLIFG 1232
Db 1419 EKCENLKKSLTR-----FAHLKQELTNKKKELTSKNAENAMQKETEESLKDSNHQLOE 1472
QY 1233 KVNIPSQSTRHSTVATECLSKNTENLILSLKNSLNDCSNOVILAKASQE-----HH 1284
Db 1473 SASSDAEQITKEQFOLKSEKERTKELADSKNEBLEHLOEAVDADGKTEISNLEKEIHE 1532
QY 1285 LGEETKCSASLFSQCSLEDLTANTINTQDPP-----LIGSKQMRHOSQGV 1333
Db 1533 LRSDKBGLVQOVQNLASAEALAREHSPHQGSLENADEIARLSQLESTKQYYEKETEI 1592
QY 1334 GLSDRELVSDDERTGTLBENNQBQSDMSNGEAAACSETSVSDCGLSSQSDILT 1393
Db 1593 LAARSELVAEKEKTBELQENQUNERSQRIKELEEQAOQRNSSENT-----HDNIDDMIK 1645
QY 1394 TQORDTMOHNL-----IKLOQEMAELE-----AVLEQHGQSPNSYPSIISDSSAL-ED 1441
Db 1646 QVVEEKLKENSANFVKUKLVVAETEFKSAKISVYEKTRDLQNKITQLEETIENLNKQ 1705
QY 1442 LRNPEQS-----TSEKAVLTSQKSEPIYSONPEGLSADKFEVSADSSTSK-----NKE 1490
Db 1706 LSNPKTDESTSVTETKPVTSKPTASKADVQGNATEASSAKREPSSGKLSARLOGTGKQ 1765
QY 1491 PGVERSSPK 1500
Db 1766 KGVRPAVSR 1775
```

```
RESULT 11
JC5837
364K Golgi complex-associated protein - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000
C:Accession: JC5837
R:Toki, C.; Fujiwara, T.; Sohma, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.
Cell Struct. Funct. 22, 565-577, 1997
A:Title: Identification and characterization of rat 364-kDa Golgi-associated protein
A:Reference number: JC5837; MUID:98093490; PMID:9431462
A:Accession: JC5837
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3187 <TOK>
A:Cross-references: DBJ:D25543; NID:g516825; PIDN:BAA05026.1; PID:g516826
C:Comment: This protein plays a role in the formation and maintenance of the characte
C:Superfamily: giantin
F:49-549,624-1176,1238-1707,1763-3114/Domain: coiled-coil leucine zipper #status pred
F:3165-3187/Domain: membrane anchor #status predicted <MAD>
```

```
Query Match 3.5%; Score 340.5; DB 2; Length 3187;
Best Local Similarity 18.9%; Pred. No. 6.2e-07;
Matches 375; Conservative 326; Mismatches 719; Indels 561; Gaps 91;
```

```
QY 23 ECPICLELIKPEVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITKSLQESTRFSQL 82
Db 919 EEPVCKEALQOELE-----WLRKESEQRKKLQAAL---ISRKELIQ--KVSKL 962
QY 83 VEELKIKITCAQFDLTGLEYANSYFNFAKKENNSPEHLKDEVSIIQSGYRNRAKRLQSEP 142
Db 963 EEELAKV-----REESTKQSL-----RESEKRELEEDS 990
QY 143 ENP-----SLQETSLSVOLSNLGTVRTLRTKQRTQPOKTSVYIELGSDS 186
Db 991 KKKDDPEKYGTSEWRELEVSLRLTISEKEVELEGIIRDLKEKAAAEELQALQVRM--- 1046
QY 187 SEDTVNKATYCSVGDOELL-QITPQ---GTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 1047 TQDLQNKTKQIDLQEEITENQATIQKFTGMDAGDGSQAVKETSVPSPPRAGGGEHWK 1106
QY 241 PSNNDLNTTKRAAERHPKPYQGSVSNLHVPCGTNTHASSLQHSNLSLLLTDKRMNVE 300
Db 1107 PE-----LEKIVDLEKEKTQ-----LQKKLQEAALISRAI-LK 1139
QY 301 KAEFCNK-----SKQGLARSOHNRWAGSKETCNDRRTPSTPEKKVVDLNADPLCERKEN 354
Db 1140 KAQEKEKHLKELKEQKDAYRHLOEQFDGQSKENENIRAPLRQ-----LQAKEST 1189
QY 355 KOKLPCSENPROTQEDVPWITLNSSIQKVNEWFSRDELLGSDSDSHDGES---ESNAKVAD 411
Db 1190 DQQLPGTGOQPTGSGESLSEGT-----EPASESD-LHAAQPSHPGETATLQATVSVAQ 1243
QY 412 VLDVLNVEYSGSEKIDLLASDPHEALICKSERV----- 447
Db 1244 IQDQLKELEV---EKEELEKISSTTSITSELTKKSEVLLLOEQINEOGLQIKNKAASHEA 1300
QY 448 --HKSVESNITEDKIFGTYRKKASLPNL-----SHVTE-----NLIIGAFVTE 489
Db 1301 KAHTEQKQOELE-----SSQLKIADLEHLKTLQPELTLQKHVGQKEEREVSVLVGLGEEK 1355
QY 490 PQITQERPLTNKLRKRPTSGLHP-----EDFIKKADLAV---QKTEMINQGTN 537
Db 1356 EOTL--TTVQTEMEBEQERLIKALHTQLEMQAKEHEERLUKQVQVEICELKKQKEEESK 1413
QY 538 QTEQNGQVNNITNSGHENKTKGDSIQNEKNPNPIESLEKE-SAFKTKAPPISSISNMEL 596
Db 1414 AKQO-----LQRKLOAALISRKALKENKSLQEQLSARDADVEHLTKSLADVES 1462
QY 597 ELNTHNSKAPKKNRLRRKSSSTRHHALEVLVSRNL-----PPNCTELQIDSSCSSSEE 649
Db 1463 QVSVQNGQ---EKDALLGKLLALQOEERDKLIVEMDKSLLENQSLGGSCESLKLALGLTLED 1519
```


Query Match				3.5%; Score 334.5; DB 2; Length 3328;			
Best Local Similarity				18.9%; Pred. No. 1.2e-06;			
Matches 399; Conservative 286; Mismatches 724; Indels 701; Gaps 96;							
QY	72	SLOESTRFSOLVELLKIIIC-----AFOLDGTGLEVANSYNFAKKENNS----	PEHLKD 120				
DB	466	SLEDISIAKGQWMTSOAALSPSIRKSIFKMRLEPLDETGLTGVFSDSMTNSTTTEEHEAS	525				
QY	121	EVSTIQSMGYRNRAKRLQLOSEPNS-----LQETSLSVQLSNLGTVTRLTKQRIQPKQT	176				
DB	526	ACGLGILTACQSRSDSCPSVDTGSWPTTLTDTSATVK--NAGLISTLKNKR-----	577				
QY	177	SVYTELGSDSEDVTKATYCSVGDQELLQITPQGTDRDEISLDSAKKAACEFSETDV----	233				
DB	578	-----KFIY-SVSDASLQKKLQTHRQLELTNL-SAQLEASAFEVPLT	619				
QY	234	-TWTEHHQPSNNDLNTTEKRAERHEPEKYQSSVSNLHVPCGNTTHASSLQHENSSL--	290				
DB	620	FTVNSGIPDSSD-----KKRCLPNDPE-----EPLSNFSGTATSKLSYVHA	663				
QY	291	LLTKDRMNVEKAEFCNKSQPLARSQHNRWAGSKETC-NDRRTPTSTPEKKVDLNAADPLCE	349				
DB	664	LISQD-LNDKEAIVIEEKPPQYTAREADFLCLPTECENDQKSPKYSN-----	711				
QY	350	RKEWNKOKLPCSENPRDTEVPWITLNSIQKYNEMFSRDELGLGSDSDHDGSESNAKV	409				
DB	712	-----GKEKVLVS-----ACLPASVOLSSISFESQENHLG---DHNGTSTLKLTP	753				
QY	410	ADVLDVLNEVDEYSGSEKIDLLASDPHEALICK-SERVHKSVSERNIEDKIFGKT----	464				
DB	754	SSKLPL-----SKADWVSREK-----MCKMPEKLQCESCKVNIE---LSKNILEV	795				
QY	465	-----YRKASPLNLSHVTENLIIGAFVTEPQLIOERPLTNKLKRRRPTSGLHPDEFIK	519				
DB	796	NEICILSENSKTPGLPPGNIIEVASMKSQFNQNAKIVIQDKQKSP-----FI-	846				
QY	520	KADLAVOKTPEMI--NOGTNOTEQNGOVMNITNSGHE-----NKTGDSIQN-----	564				
DB	847	-SEVAVNMNSEELFPDGSNFAFQVTKCNKPDGLSSVELQEEEDLSHTQGPSLKNSPMAV	905				
QY	565	-----EKNPNPIESLEK-----ESAFKTKA-----	584				
DB	906	DEDVDDAHAAQVILITKDSLSLAVVHDYTEKSRNNIEHQHQGTEDKDFKSNLSNLMKSDGN	965				
QY	585	-----EPISS-----SISNMELELNINHSKAPKK--NRLRRKSSTR-----	618				
DB	966	SDCSKDWSEFLDPVLNHNFCGSPRTASNKEIKLSEHNVKSKMFFKDIEQYPTRLACID	1025				
QY	619	HIHALELVSRNLSPPNCTELQIDSCSSSEIEKKKYQNMPVHRHSRNQLMEGKEPATGA	678				
DB	1026	IVNTLPLANGKKLSEPHIFDLK-----SVTTVSTQSHNQSSVSH-----	1064				
QY	679	KSNKKNQEQTSKRHSDTPELKLITNAPGSFTKCSNTSELKEFVNPSLPREEKEKLETV	738				
DB	1065	-----EDTDAPQM-LSSKQDFHSNNLTTSQKAEITELSTILEESGSOFEFT	1110				
QY	739	K-----VSNNAEDPKDMLSGERVLOT-ERSVESSISLVPGTDYGTQESISLLEVST	790				
DB	1111	QFRKPSHIAQNTSEVPGNQW-----VLSTASKEWKOTDHLPLVDPDSVGQTDHKSQFEGSA	1166				
QY	791	LGAKEPTNPKCVSQCAAFENPKGLIHGCSKDRNNDTEGFKY--PLGHEVNHNSRSTSIEME	848				
DB	1167	-GVKQSPFHLLDTCN-KNT-----SCFLPNINEMEFGRICSAIGTKLSVSN-----	1212				
QY	849	ESELDQAYLQNTFKVKRQSAFALFSPNGNAEBECA-----TFSAHSGS-----LKKQ	895				
DB	1213	-----ALRKAMKLFSDIENSEBPSAKVGRPGFSSSAHSDHSVASVFKIKKP	1257				
QY	896	SPKVTE-----CEQKEENQ-----GKN-ESNIKPVQ-----	T 922				
DB	1258	NTEKSEDEKSSKQVTLQNNIEMTTCIFVGRNPEKVIKNTKHEDSYTSSORNNLENSDGS	1317				

QY	923	VNITAGFPVV---GQKDKPVD-NAKCSIKGSGRFLCSSQFRGNETGLITPNKHLQNPY	978
DB	1318	MSSTSG-PVYIHKGDSLDLPADGSKC-----PESC--TQVAREENTQIKENISDLTCLFI	1369
QY	979	RIPPLPIKSFVTKCKKNLLLEENFEHSHSPEREMGNENIPSTVSTISRNNIRNVFXE	1038
DB	1370	MKAETCMKSGSDRKQLPSPDKMERNIKEFNIS-----FOTASGKNTR--VSKE	1414
QY	1039	ASSNSINEVGSSSTNEIGSSDENEIQAELGRNRGPKLNAMLRLGLVQLQPEYKQSL	1098
DB	1415	SLKSNVINFRNDEL-----TWISLNSKILHGINKDKMHTSCHKKAISIKKVFEDHF	1469
QY	1099	PGSNCKH-PEIKKOEYEVQTVNTDFSPYLIS-----DNLEQPMGSSHASQ	1144
DB	1470	PIVTVSOLPAQOQPEVE-----TESTKEPTLLSFHTASGKKVKIMQESLDK-----	1518
QY	1145	VCSETPDDLLDDEGEIREDTSAEND--IKESSAVFSKSVQKGELSRSPPFTTHLAQGY	1202
DB	1519	LCGETQ-----YVRKTASFQSGSKPLKDSKKELTAYEKIEVT-----	1556
QY	1203	RRGAKKLESSEENLSSEDEELPCFQHLFGKVNNPISQSTRHSTVATE-----	1250
DB	1557	---ASKCEMNQNFVSKETEMLPQOONYHMYWOTENLKTNGTSSKVOENIENNVKPNRIW	1613
QY	1251	CLSKN-----TEENLLS--LKNSLNDCSNQVILAK-----	1278
DB	1614	CICQSSYPVTEDSALAYTTEDSRKTCVRESSLSKGRKWLREOGDKLGTNRNTKIECVKEH	1673
QY	1279	-----ASEPHH-----LSEETHKCSASLFSQCSSELDL----	1306
DB	1674	TEDFAGNASYEHLVLIIRTEIDTNHVSNOVSTLLSDPNVCHSYLQSQSFCHCDMHNDS	1733
QY	1307	-----TAN-----TNTQ-----	1313
DB	1734	GYFLKNKIDSDVPPDMKNABGNITSPRVATKERNLHPQITINEYCVOKLETINTSPHANKD	1793
QY	1314	---DPFLIGSSK-----QMRHQSBSQGVGLSDKELVSDDEERTGTGLENNQEQSMD	1362
DB	1794	VAIDPSLLDSNRNCKVGLSVFITAHSQETE-----RTKEIVTDNCYKIV-----EQNRQ	1841
QY	1363	SNLGEAASGCESETSVSDCSGLSSQSDILTTQOROT-----MQHNLKILQOEMAE	1414
DB	1842	SKPDTQCTSCHVKLDDSDKDFICPSSGDGVCINSRKDSFCLHNEQILQHN-----QSMS--	1894
QY	1415	EAVLEHQGPSNPSYPSIISDSSALEDLNRPEO-----STSEKAVLTSOKSSE	1462
DB	1895	-----GLKKAATPPVGLTWTDSKISIREPPQAAHPRTYGFSTASGKAIQVSDASLE	1947
QY	1463	YPIQNPEGLSADKFEVSADSTSKNKP--GVERSPPSKCPSLDDRWYMHSCSGLQ--	1518
DB	1948	-KARQVFSEMDGDAKQLSSMVSLGNEKPHHSVKRENS-----VVHSTQGVLSLP	1996
QY	1519	-----NRNP-----SQEELIKVVDV-----EQOOLEESG-PHDLTE	1549
DB	1997	KPLPGMVNSSVFGSFTAGGKLVTVSESALHKVKGMLLEEFDLIRTEHTLQHSPIPEDVSK	2056
QY	1550	TSYLPQDQLEGTPY-LESGISLFDPSDPSDRAPEASARVGNIPSTTSALAKVPLKVA	1608
DB	2057	ILPQCAEINTPEYPVNSKLQKTYNDKSLPSNYK--ESGSSGNTOSIEVLSLQSMERN	2114
QY	1609	ESAQSPAAAHTTDTAGYNAMESVSREKPELTASTERVNKRMSM-----VVSGLTPEEF	1662
DB	2115	QDTQL-----VLGTVKSHSKANLLGKEQTLPKNIKVKTDKMTFSDVPVKTN	2161
QY	1663	MLVYKFKARKHIHILTLNITEETHVVNMK-----TDAE-----FVCERTKLYFLGIAGG	1710
DB	2162	VGEYYSKESENYFETEAV--ESAKAFMEDELDTSEQTHAKCSLFTCPQNETLFSNRTRK	2219
QY	1711	KWVVSFWTWTQSIKERKMLNEHD--FEVRGDVYNGRNHQCPKARES---QDRKIF---R	1762
DB	2220	RGVTVDAVGQPIKRIKSLLENEFDRIIESKGSALT-----PSKSTPDGTVKDRSLRTHHM	2273
QY	1763	GLEICCYGPF	1772

Db 2274 SLEPVTGPF 2283

Search completed: January 22, 2003, 16:52:06
Job time : 96.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 16:43:17 ; Search time 17.5 Seconds
(without alignments)
4415.451 Million cell updates/sec

Title: US-09-734-672-4
Perfect score: 9649
Sequence: 1 MDLSALRVEEVONVINAMOK.....LYQCQLDYLIPQIPHSY 1863

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9642	99.9	1863	1 BRC1_HUMAN	P38398 homo sapien
2	6923	71.7	1878	1 BRC1_CANFA	Q95153 canis famli
3	4817	49.9	1812	1 BRC1_MOUSE	P48754 mus musculus
4	357.5	3.7	3329	1 BRC2_MOUSE	P97929 mus musculus
5	340	3.5	2869	1 RBP1_PLAVB	Q00798 plasmodium
6	339	3.5	1957	1 YD86_SCHPO	Q10411 schizosacch
7	327	3.4	1875	1 MLP1_YEAST	Q02455 saccharomyc
8	327	3.4	3418	1 BRC2_HUMAN	P51587 homo sapien
9	325	3.4	3911	1 AKA9_HUMAN	Q99996 h a-kinase
10	323.5	3.4	3210	1 CENF_HUMAN	P49454 homo sapien
11	322	3.3	1658	1 YM67_YEAST	Q03661 saccharomyc
12	321.5	3.3	4377	1 ANK3_HUMAN	Q12955 homo sapien
13	319.5	3.3	2492	1 ATRX_HUMAN	P46100 homo sapien
14	317.5	3.3	1972	1 P531_HUMAN	Q12888 homo sapien
15	316	3.3	1727	1 ALM1_SCHPO	Q90457 schizosacch
16	313.5	3.2	1679	1 YI09_YEAST	P40457 saccharomyc
17	312	3.2	3924	1 ANK2_HUMAN	Q01484 homo sapien
18	310.5	3.2	2845	1 APC_MOUSE	Q61315 mus musculus
19	308.5	3.2	2748	1 NUM1_YEAST	Q00402 saccharomyc
20	308	3.2	2230	1 GOG4_HUMAN	Q13439 homo sapien
21	307.5	3.2	1790	1 US01_YEAST	P25386 saccharomyc
22	304.5	3.2	1744	1 TANA_XENLA	Q01550 xenopus lae
23	303.5	3.1	1539	1 Y373_HUMAN	Q15078 homo sapien
24	301.5	3.1	2319	1 AKA6_HUMAN	Q13023 homo sapien
25	301	3.1	2464	1 MAPB_MOUSE	P14873 mus musculus
26	299.5	3.1	2843	1 APC_HUMAN	P25054 homo sapien
27	296.5	3.1	1940	1 MYH3_HUMAN	P11055 homo sapien
28	294.5	3.1	2349	1 TPR_HUMAN	P12270 homo sapien
29	293	3.0	2468	1 MAPB_HUMAN	P46821 homo sapien
30	290	3.0	2663	1 CENF_HUMAN	Q02224 homo sapien
31	290	3.0	3562	1 PGCV_CHICK	Q90953 gallus gall
32	288.5	3.0	1928	1 MYL1_YEAST	P08964 saccharomyc
33	288	3.0	2476	1 ATRX_MOUSE	Q61687 mus musculus

RESULT 1				
BRC1_HUMAN				
ID	BRC1_HUMAN	STANDARD;	PRT;	1863 AA.
AC	P38398;			
DT	01-OCT-1994 (Rel. 30, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Breast cancer type 1 susceptibility protein.			
GN	BRCAL.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND VARIANT ARG-1775.			
RX	MEDLINE=95025896; PubMed=7545954;			
RA	Miki Y., Svensen J., Shattuck-Eidens D., Futreal P.A., Harshman K.,			
RA	Tavtigian S., Liu Q., Cochran C., Bennett L.M., Ding W., Bell R.,			
RA	Rosenthal J., Hussey C., Tran T., McClure M., Frye C., Hattier T.,			
RA	Phelps R., Haugen-Strano A., Katcher H., Yakumo K., Gholami Z.,			
RA	Shaffer D., Stone S., Bayer S., Wray C., Bogden R., Dayananth P.,			
RA	Ward J., Tonin P., Narod S., Bristow P.K., Norris F.H., Helvering L.,			
RA	Morrison P., Rostock P., Lai M., Barrett J.C., Lewis C., Neuhausen S.,			
RA	Cannon-Albright L., Godiglar D., Wiseman R., Kamb A., Skolnick M.H.;			
RT	"A strong candidate for the breast and ovarian cancer susceptibility			
RT	gene BRCAL.";			
RL	Science 266:66-71(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97092865; PubMed=8938427;			
RA	Smith T.M., Lee M.K., Szabo C.I., Jerome N., McEuen M., Taylor M.,			
RA	Hood L., King M.-C.;			
RT	"Complete genomic sequence and analysis of 117 kb of human DNA			
RT	containing the gene BRCAL.";			
RL	Genome Res. 6:1029-1049(1996).			
RN	[3]			
RP	IDENTIFICATION OF BRCAL AS MEMBER OF BASC.			
RX	MEDLINE=20245492; PubMed=10783165;			
RA	Wang Y., Cortez D., Yazdi P., Neff N., Elledge S.J., Qin J.;			
RT	"BASC, a super complex of BRCAL-associated proteins involved in the			
RT	recognition and repair of aberrant DNA structures.";			
RL	Genes Dev. 14:927-939(2000).			
RN	[4]			
RP	REVIEW ON VARIANTS.			
RX	MEDLINE=96400954; PubMed=8807330;			
RA	Couch F.J., Weber B.L.;			
RT	"Mutations and polymorphisms in the familial early-onset breast			
RT	cancer (BRCAL) gene." Breast Cancer Information Core.";			
RL	Hum. Mutat. 8:18-18(1996).			
RN	[5]			
RP	VARIANTS LEU-1637; GLU-1708 AND ARG-1775.			
RX	MEDLINE=95025878; PubMed=7939630;			
RA	Futreal P.A., Liu Q., Shattuck-Eidens D., Cochran C., Harshman K.,			
RA	Tavtigian S., Bennett L.M., Haugen-Strano A., Svensen J., Miki Y.,			
RA	Eddington K., McClure M., Frye C., Weaver-Felhaus J., Ding W.,			
RA	Gholami Z., Soedervekist P., Terry L., Jhanwar S., Berchuk A.,			

RA Iglehart J.D., Marks J., Ballinger D.G., Barrett J.C., Skolnick M.H.,
 RA "BRCA1 mutations in primary breast and ovarian carcinomas.,"
 RT Science 266:120-122(1994).
 RL [6]
 RN
 RP VARIANTS BC GLY-64, AND VARIANTS ALA-772; ASN-1040 AND GLY-1443.
 RX MEDLINE-95201806; PubMed=7894491;
 RA Castilla L.H., Couch F.J., Erdos M.R., Hoskins K.F., Calzone K.,
 RA Garber J.E., Boyd J., Lubin M.B., Deshano M.L., Brody L.C.,
 RA Collins F.S., Weber B.L.;
 RT "Mutations in the BRCA1 gene in families with early-onset breast and
 RT ovarian cancer.,"
 RT Nat. Genet. 8:387-391(1994).
 RL [7]
 RN
 RP VARIANT BC G-61, AND VARIANTS R-356; G-1038; N-1040; R-1183 AND
 RP G-1613.
 RX MEDLINE-95201808; PubMed=7894493;
 RA Friedman L.S., Ostermeyer E.A., Szabo C.I., Dowd P., Lynch E.D.,
 RA Rowell S.E., King M.-C.;
 RT "Confirmation of BRCA1 by analysis of germline mutations linked to
 RT breast and ovarian cancer in ten families.,"
 RL Nat. Genet. 8:399-404(1994).
 RL [8]
 RN
 RP VARIANT BC GLY-61.
 RX MEDLINE-96108965; PubMed=8554067;
 RA Serova O., Montagna M., Torchard D., Narod S.A., Tonin P., Sylla B.,
 RA Lynch H.T., Feunteun J., Lenoir G.M.;
 RT "A high incidence of BRCA1 mutations in 20 breast-ovarian cancer
 RT families.,"
 RL Am. J. Hum. Genet. 58:42-51(1996).
 RL [9]
 RN
 RP VARIANT BOC TRP-841.
 RX MEDLINE-97123469; PubMed=8968716;
 RA Barker D.F., Almeida E.F.A., Casey G., Fain P.R., Liao S.-Y.,
 RA Masunaka I., Noble B., Kurosaki T., Anton-Culver H.;
 RT "BRCA1 R841W: a strong candidate for a common mutation with moderate
 RT phenotype.,"
 RL Genet. Epidemiol. 13:595-604(1996).
 RL [10]
 RN
 RP VARIANTS BC AND BOC.
 RX MEDLINE-96372821; PubMed=8776600;
 RA Durocher F., Shattuck-Eidens D., McClure M., Labrie F.,
 RA Skolnick M.H., Goldgar D.E., Simard J.;
 RT "Comparison of BRCA1 polymorphisms, rare sequence variants and/or
 RT missense mutations in unaffected and breast/ovarian cancer
 RT populations.,"
 RL Hum. Mol. Genet. 5:835-842(1996).
 RL [11]
 RN
 RP VARIANTS BC MET-271 AND SER-1150.
 RX MEDLINE-96303704; PubMed=8723683;
 RA Katagiri T., Emi M., Ito I., Kobayashi K., Yoshimoto M., Iwase T.,
 RA Kasumi F., Miki Y., Skolnick M.H., Nakamura Y.;
 RT "Mutations in the BRCA1 gene in Japanese breast cancer patients.,"
 RL Hum. Mutat. 7:334-339(1996).
 RL [12]
 RN
 RP VARIANT BC GLY-61, AND VARIANTS ARG-239; TRP-841 AND ILE-1512.
 RX MEDLINE-98430998; PubMed=9760198;
 RA Dong J., Chang-Claude J., Wu Y., Schumacher V., Debatin I., Tonin P.,
 RA Royer-Pokora B.;
 RT "A high proportion of mutations in the BRCA1 gene in German
 RT breast/ovarian cancer families with clustering of mutations in the 3'
 RT third of the gene.,"
 RL Hum. Genet. 103:154-161(1998).
 RL [13]
 RN
 RP VARIANT BC GLY-64, AND VARIANTS.
 RX MEDLINE-98141685; PubMed=9482581;
 RA Andersen T.I., Eiken H.G., Couch F., Kaada G., Skrede M., Johnsen H.,
 RA Aloysius T.A., Tveit K.M., Tranebjaerg L., Doerum A., Moeller P.,
 RA Weber B.L., Boerresen-Dale A.-L.;
 RT "Constant denaturant gel electrophoresis (CDGE) in BRCA1 mutation
 RT screening.,"
 RL Hum. Mutat. 11:166-174(1998).
 RL [14]
 RN
 RP VARIANTS BC S-22; L-461; D-465; V-552; S-892; D-960; I-1025 AND
 RP A-1047.
 RX MEDLINE-98272917; PubMed=9609997;
 RA Katagiri T., Kasumi F., Yoshimoto M., Nomizu T., Asaishi K., Abe R.,
 RA Tsuchiya A., Sugano M., Takai S., Yoneda M., Fukutomi T., Namba K.,
 RA Makita M., Okazaki H., Hirata K., Okazaki M., Furutsuma Y.,
 RA Morishita Y., Iino Y., Karino T., Ayabe H., Hara S., Kajiwara T.,
 RA Hougou S., Shimizu T., Toda M., Yamazaki Y., Uchida T., Kunitomo K.,
 RA Sonoo H., Kurebayashi J.-I., Shimotsu K., Nakamura Y., Miki Y.;
 RT "High proportion of missense mutations of the BRCA1 and BRCA2 genes in
 RT Japanese breast cancer families.,"
 RL J. Hum. Genet. 43:42-48(1998).
 RL [15]
 RN
 RP VARIANT OC ARG-1749.
 RX MEDLINE-20455732; PubMed=10486320;
 RA Gayther S.A., Russell P., Harrington P., Antoniou A.C., Easton D.F.,
 RA Ponder B.A.J.;
 RT "The contribution of germline BRCA1 and BRCA2 mutations to familial
 RT ovarian cancer: no evidence for other ovarian cancer-susceptibility
 RT genes.,"
 RL Am. J. Hum. Genet. 65:1021-1029(1999).
 RL [16]
 RN
 RP VARIANT BC S-346, AND VARIANTS L-871; G-1038; R-1183 AND G-1613.
 RX MEDLINE-99254821; PubMed=10323242;
 RA Li S.-L., Tseng H.-M., Yang T.-P., Liu C.-H., Teng S.-J.,
 RA Huang H.-W., Chen L.-M., Kao H.-W., Chen J.H., Tseng J.-H., Chen A.,
 RA Hou M.-F., Huang T.-J., Chang H.-T., Mok K.-T., Tsai J.-H.;
 RT "Molecular characterization of germline mutations in the BRCA1 and
 RT BRCA2 genes from breast cancer families in Taiwan.,"
 RL Hum. Genet. 104:201-204(1999).
 RL [17]
 RN
 RP VARIANTS OC, AND VARIANTS.
 RX MEDLINE-99214030; PubMed=10196379;
 RA Janezic S.A., Zlogas A., Krumroy L.M., Krasner M., Plummer S.J.,
 RA Cohen P., Glidea M., Barker D., Haile R., Casey G., Anton-Culver H.;
 RT "Germline BRCA1 alterations in a population-based series of ovarian
 RT cancer cases.,"
 RL Hum. Mol. Genet. 8:889-897(1999).
 RL [-] FUNCTION: NOT KNOWN. MAY REGULATE GENE EXPRESSION. INVOLVED IN
 CC TRANSCRIPTIONAL REGULATION OF P21 IN RESPONSE TO DNA DAMAGE.
 CC [-] SUBUNIT: PART OF THE BRCA1-ASSOCIATED GENOME SURVEILLANCE COMPLEX
 CC (BASC), WHICH CONTAINS BRCA1, MSH2, MSH6, MLH1, ATM, BLM, PMS2 AND
 CC THE RAD50-MRE11-NBS1 PROTEIN COMPLEX. THIS ASSOCIATION COULD BE A
 CC DYNAMIC PROCESS CHANGING THROUGHOUT THE CELL CYCLE AND WITHIN
 CC SUBNUCLEAR DOMAINS. CTIP INTERACTS SPECIFICALLY WITH THE BRCT
 CC DOMAINS.
 CC [-] SUBCELLULAR LOCATION: Nuclear.
 CC [-] POLYMORPHISM: THERE IS EVIDENCE THAT THE PRESENCE OF THE RARE FORM
 CC OF Q356R AND L871P POLYMORPHISMS MAY BE ASSOCIATED WITH AN
 CC INCREASED RISK FOR DEVELOPING OVARIAN CANCER.
 CC [-] DISEASE: BREAST CANCER (BC) IS AN EXTREMELY COMMON MALIGNANCY,
 CC AFFECTING ONE IN EIGHT WOMEN DURING THEIR LIFETIME. A POSITIVE
 CC FAMILY HISTORY HAS BEEN IDENTIFIED AS MAJOR CONTRIBUTOR TO RISK OF
 CC DEVELOPMENT OF THE DISEASE, AND THIS LINK IS STRIKING FOR EARLY-
 CC ONSET BREAST CANCER. MUTATIONS IN BRCA1 ARE THOUGHT TO BE
 CC RESPONSIBLE FOR 45% OF INHERITED BREAST CANCER AND MORE THAN 80%
 CC OF INHERITED BREAST AND OVARIAN CANCER (BOC). MOREOVER, BRCA1
 CC CARRIERS HAVE A 4-FOLD INCREASED RISK OF COLON CANCER, WHEREAS
 CC MALE CARRIERS FACE A 3-FOLD INCREASED RISK OF PROSTATE CANCER.
 CC [-] SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 CC [-] SIMILARITY: CONTAINS 2 BRCT DOMAINS.
 CC [-] DATABASE: NAME=HotMolBase; NOTE=BRCA1 entry;
 CC WWW="http://bioinformatics.weizmann.ac.il/hotmolebase/entries/brca1.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U14680; AAA73985.1; -.

Query Match									
Best Local Similarity 99.9%; Score 9642; DB 1; Length 1863;									
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
QY	1	MDLSALRVEVQVNIAMQKILECPTCLELILKEPVTCKDHIFCKFCMKLLNKKQKGPSQ	60						
DB	1	MDLSALRVEVQVNIAMQKILECPTCLELILKEPVTCKDHIFCKFCMKLLNKKQKGPSQ	60						
QY	61	CPCKNDITKRSLOESTRFSOLVEELLKIICAFQDGTGLEAYANSYFPAKKENNSPEHLKD	120						
DB	61	CPCKNDITKRSLOESTRFSOLVEELLKIICAFQDGTGLEAYANSYFPAKKENNSPEHLKD	120						
QY	121	EVSIIQSMGYRNRAKRLLOSEPENPSLQETSLSVQLSNLGTVRTLTKRIQPKTSVYI	180						
DB	121	EVSIIQSMGYRNRAKRLLOSEPENPSLQETSLSVQLSNLGTVRTLTKRIQPKTSVYI	180						
QY	181	ELGSDSEEDTVNKATYCSVGDDQLLOITPQGTREDEISLSAKKAACEFSETDVTNTEHHQ	240						
DB	181	ELGSDSEEDTVNKATYCSVGDDQLLOITPQGTREDEISLSAKKAACEFSETDVTNTEHHQ	240						
QY	241	PSNNDLNTTEKRAAERHPEKYQGSSVSNLHVPCGTNTHASSLQHENSLLLTCKDRMNYE	300						
DB	241	PSNNDLNTTEKRAAERHPEKYQGSSVSNLHVPCGTNTHASSLQHENSLLLTCKDRMNYE	300						
QY	301	KAFCNKSQOPGLARSOHNRWAGSKETCNDRRTPSTPEKKVLDNADPLCKERKWNKQKLP	360						
DB	301	KAFCNKSQOPGLARSOHNRWAGSKETCNDRRTPSTPEKKVLDNADPLCKERKWNKQKLP	360						
QY	361	SENPRTEDPWITLANSQKYNWFVSRDELGSDSDHGDGESENAKVADVLVDLNEVD	420						
DB	361	SENPRTEDPWITLANSQKYNWFVSRDELGSDSDHGDGESENAKVADVLVDLNEVD	420						
QY	421	EYSGSSEKIDLLASDPHEALICKSERVHKSVESNTEKIFGKTYRKKASLPNLSHVTE	480						
DB	421	EYSGSSEKIDLLASDPHEALICKSERVHKSVESNTEKIFGKTYRKKASLPNLSHVTE	480						
QY	481	LIITGAVTEPQIIQERPLTNKLRKRRTPSGLHPEDFIKKADLAVOKTPEMINQGTNQE	540						
DB	481	LIITGAVTEPQIIQERPLTNKLRKRRTPSGLHPEDFIKKADLAVOKTPEMINQGTNQE	540						
QY	541	QNGOVNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPDISSISNMELELNI	600						
DB	541	QNGOVNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPDISSISNMELELNI	600						
QY	601	HNSKAPKKNRLRRKSTRIHALVVRNLSPPNCTELQIDSCSSSEIKKKKYNOMP	660						
DB	601	HNSKAPKKNRLRRKSTRIHALVVRNLSPPNCTELQIDSCSSSEIKKKKYNOMP	660						
QY	661	RHSRNLQMEGKEPATGAKSKNPNQTSKRHSDPTFPELKLTPNAPGFTKCSNTSELKE	720						
DB	661	RHSRNLQMEGKEPATGAKSKNPNQTSKRHSDPTFPELKLTPNAPGFTKCSNTSELKE	720						
QY	721	FYNPSLPREEKEKLETVKVNNAEDPKDMLSGERVLOTERSVESSSIISLVPDGTGYTQ	780						
DB	721	FYNPSLPREEKEKLETVKVNNAEDPKDMLSGERVLOTERSVESSSIISLVPDGTGYTQ	780						
QY	781	ESISLLEVTSLGKAKTEPNKVSQCAAFENPKGLIHGCSKDRNDTEGPKYPLGHEVNH	840						
DB	781	ESISLLEVTSLGKAKTEPNKVSQCAAFENPKGLIHGCSKDRNDTEGPKYPLGHEVNH	840						
QY	841	RETSIEMEESLDAQYLQNTFKVSRQSFALFSPNGNAEEECATFFSAHSGSLKKQSPKVT	900						
DB	841	RETSIEMEESLDAQYLQNTFKVSRQSFALFSPNGNAEEECATFFSAHSGSLKKQSPKVT	900						
QY	901	FECQEEHQGNESNIKPVQVNTAGFPVVGQDKPVDNAKCSITKGSRCFLSSQFRG	960						
DB	901	FECQEEHQGNESNIKPVQVNTAGFPVVGQDKPVDNAKCSITKGSRCFLSSQFRG	960						
QY	961	NETGLITPNKHGLQNPYRIPPLFPTKSVKTKCKKNLLEENFEHSMSPEREMGNENIP	1020						
DB	961	NETGLITPNKHGLQNPYRIPPLFPTKSVKTKCKKNLLEENFEHSMSPEREMGNENIP	1020						

RESULT 2			
BCL1_CANFA			
ID	BCL1_CANFA	STANDARD;	PRT; 1878 AA.
AC	Q95153;		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Breast cancer type 1 susceptibility protein homolog.		
GN	BRCA1.		
OS	Canis familiaris (Dog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		

NCBI_TaxID=9615;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=97026289; PubMed=8872468;
 RA Szabo C.I., Wagner L.A., Francisco L.V., Roach J.C., Argonza R.,
 RA King M.-C., Ostrander E.A.;
 RT "Human, canine and murine BRCA1 genes: sequence comparison among
 RT species.";
 RL Hum. Mol. Genet. 5:1289-1298(1996).
 CC -!- FUNCTION: NOT KNOWN. MAY REGULATE GENE EXPRESSION. INVOLVED IN
 CC TRANSCRIPTIONAL REGULATION OF P21 IN RESPONSE TO DNA DAMAGE (BY
 CC SIMILARITY).
 CC -!- SUBUNIT: CTIP INTERACTS SPECIFICALLY WITH THE BRCT DOMAINS (BY
 CC SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 CC -!- SIMILARITY: CONTAINS 2 BRCT DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U50709; AAC48663.1; -.
 DR InterPro: IPR001357; BRCT.
 DR InterPro: IPR002378; Brst_cancer1.
 DR Pfam: PF00097; Znf_C3HC4; 1.
 DR Pfam: PF00533; BRCT; 2.
 DR PRINTS: PR00493; BRSTCANCER1.
 DR SMART: SM00292; BRCT; 2.
 DR SMART: SM00184; RING; 1.
 DR PROSITE: PS00172; BRCT; 2.
 DR PROSITE: PS00518; ZF_RING_1; 1.
 DR PROSITE: PS00089; ZF_RING_2; 1.
 KW Zinc-finger; DNA-binding; Nuclear protein; Anti-oncogene; Repeat.
 FT ZN-FING 24 65 RING-TYPE.
 FT DOMAIN 1652 1739 BRCT 1.
 FT DOMAIN 1764 1863 BRCT 2.
 SQ SEQUENCE 1878 AA; 208446 MW; 4799B650327C4F CRC64;

 Query Match 71.7%; Score 6923; DB 1: Length 1878;
 Best Local Similarity 74.2%; Pred No. 2.5e-282;
 Matches 1390; Conservative 165; Mismatches 301; Indels 18; Gaps 13;

 QY 1 MDLSALRVEEVQVNVINAMOKILECPICELIKEPVTSCDHIFCKFCMLKLLNKGPSQ 60
 DB 1 MDLSADRVVEEVQVNVINAMOKILECPICELIKEPVTSCDHIFCKFCMLKLLNKGPSQ 60

 QY 61 CPLKNDITKRSLOESTRFSQVSVLVELIKIICAFOLDTGLEYANSYNFAKKENNSPEHLKD 120
 DB 61 CPLKNDITKRSLOESTRFSQVSVLVELIKIICAFOLDTGLEYANSYNFAKKENNSPEHLKE 120

 QY 121 EYSIIQSGMYRNRRKRLQSPENPSLQETSLSVOLNGLGVRLTKORTQPOKTSVYI 180
 DB 121 EYSIIQSGMYRNRRKRLQSPENPSLQETSLSVOLNGLGVRLTKORTQPOKTSVYI 179

 QY 181 ELGSDSSEDVTNKATYCSVGQDQLQITPQGTDRDEISLDSAKKAACFESETDVTNTEHHQ 240
 DB 180 ELGSDSSEDVTNKATYCSVGQDQLQITPQGTDRDEISLDSAKKAACFESETDVTNTEHHQ 237

 QY 241 PSNNDLNTTKRAERHPKTYQGSVSNLHVPCGCTVTHASSLOHENSLLLTKDRNVE 300
 DB 241 PSNNDLNTTKRAERHPKTYQGSVSNLHVPCGCTVTHASSLOHENSLLLTKDRNVE 297

 QY 301 KAEFCNKSQKPLARSOHNRWAGSKETCNDRPTSTKPKVDLADPLCERKWNKOKLPC 360
 DB 301 KAEFCNKSQKPLARSOHNRWAGSKETCNDRPTSTKPKVDLADPLCERKWNKOKLPC 357

 QY 361 SENPRDTEVPWITLNSIRKVNWFSRSDDEILTSDSHDRGSELNTEVGGAVEVPNEVG 417
 DB 361 SENPRDTEVPWITLNSIRKVNWFSRSDDEILTSDSHDRGSELNTEVGGAVEVPNEVG 417

 QY 421 EYSSSKIKDILLASDPHEALICKSERVHKSQSVESNIEDKIFGKTYRKASLPNLSHVTEN 480
 DB 421 EYSSSKIKDILLASDPHEALICKSERVHKSQSVESNIEDKIFGKTYRKASLPNLSHVTEN 477

 QY 481 LIICAFYTEPOIIQERPLTNKLRKRRTSGLHPEDFIKKADLA-VOKTPEMINGOTNQT 539
 DB 481 LIICAFYTEPOIIQERPLTNKLRKRRTSGLHPEDFIKKADLA-VOKTPEMINGOTNQT 537

 QY 540 EQNQVNMNITSGHENKTGDSIQNEKNPNPISLEKESAFKTKAEPISISSINMELELN 599
 DB 540 EQNQVNMNITSGHENKTGDSIQNEKNPNPISLEKESAFKTKAEPISISSINMELELN 597

 QY 600 IHNSKAPKRLRLRKSSTRHIALELVVSRNLSPNCTELQIDSCSSSEETKKKYNOMP 659
 DB 600 IHNSKAPKRLRLRKSSTRHIALELVVSRNLSPNCTELQIDSCSSSEETKKKYNOMP 657

 QY 660 VRHSRLQLMEGKEPATGAKSKNKPNEQTSKRHSDTFPELKLTNAPGSETKCSNTSELK 719
 DB 660 VRHSRLQLMEGKEPATGAKSKNKPNEQTSKRHSDTFPELKLTNAPGSETKCSNTSELK 717

 QY 720 EFNPSLPREEKEKLETVKYSNNAEDPKDMLSGERVLOQTERSVESSISLVPGDTYGT 779
 DB 720 EFNPSLPREEKEKLETVKYSNNAEDPKDMLSGERVLOQTERSVESSISLVPGDTYGT 777

 QY 780 QESISLLEVTGLKAKTEPNKCVSQAAPENPKGLIHGCSKDNDRNDEGFKYPLGHEVNH 839
 DB 780 QESISLLEVTGLKAKTEPNKCVSQAAPENPKGLIHGCSKDNDRNDEGFKYPLGHEVNH 837

 QY 840 SRETSIEMEESLDAOYLONTFKVSKQSFPALFSPNGNABEECATFSAHSGSLKKQSPKV 899
 DB 840 SRETSIEMEESLDAOYLONTFKVSKQSFPALFSPNGNABEECATFSAHSGSLKKQSPKV 897

 QY 900 TFECEKEENQGNESNIKPVQTVNITAGFPVVGQK-DKPYVDNAKCSIKGSRFCLSSQF 958
 DB 900 TFECEKEENQGNESNIKPVQTVNITAGFPVVGQK-DKPYVDNAKCSIKGSRFCLSSQF 957

 QY 959 RGNETGLITPNKGLLONPYRIPPLFKIPKVFVKTKCKNLLNEENFEHSMSPERMGNEN 1018
 DB 959 RGNETGLITPNKGLLONPYRIPPLFKIPKVFVKTKCKNLLNEENFEHSMSPERMGNEN 1017

 QY 1019 -IPSTVTISRNIRENVKFEASSNINEVSGSSNINEVSGSSNINEVSGSSNINEVSGSSN 1077
 DB 1019 -IPSTVTISRNIRENVKFEASSNINEVSGSSNINEVSGSSNINEVSGSSNINEVSGSSN 1077

 QY 1078 PKNLAMRLGLVQPEVYKQSLPGSNCKKHPEIKKO-EYEEVVQTVNTDFSPYLSIDNLEOP 1136
 DB 1078 PKNLAMRLGLVQPEVYKQSLPGSNCKKHPEIKKO-EYEEVVQTVNTDFSPYLSIDNLEOP 1137

 QY 1137 MGSASHASQVCSPTDDLLDDGEIKEDTSFAENDIKESAVPSKSVQKGLSRSPSPFTH 1196
 DB 1137 MGSASHASQVCSPTDDLLDDGEIKEDTSFAENDIKESAVPSKSVQKGLSRSPSPFTH 1197

 QY 1197 HLAQGVRRGAKKLESSEENLSEDEELPCFQHLFGVNNIPSTOSTRHSVTATECLSKNT 1256
 DB 1197 HLAQGVRRGAKKLESSEENLSEDEELPCFQHLFGVNNIPSTOSTRHSVTATECLSKNT 1257

 QY 1257 EENLLSLKNSLNCNOVILAKAQEHHLSEETKCSASFSSQCSLELDLTANTNTQDPF 1316
 DB 1257 EENLLSLKNSLNCNOVILAKAQEHHLSEETKCSASFSSQCSLELDLTANTNTQDPF 1317

 QY 1317 LI--GSKQWRHSESQCVGLSDKELYS--DDEERGTLGLEN--OEEQSMDSNLGEAASGC 1372
 DB 1317 LI--GSKQWRHSESQCVGLSDKELYS--DDEERGTLGLEN--OEEQSMDSNLGEAASGC 1372

 QY 1373 ESETSVSDCSGLSSQSDILTTQORDTMQHLIKLQOEAELEAVLQHGSGOPSNVPSI 1432
 DB 1373 ESETSVSDCSGLSSQSDILTTQORDTMQHLIKLQOEAELEAVLQHGSGOPSNVPSI 1436

 QY 1433 ISDSSALEDLRNPEQSTSEKAVLTQSKSEPYISQNPPEGLSADKFEYSADSTSKNKP 1492
 DB 1433 ISDSSALEDLRNPEQSTSEKAVLTQSKSEPYISQNPPEGLSADKFEYSADSTSKNKP 1492

Db 1437 IADSCSPEDLLNPEQNASER-VITSEKSSDSPISQNPESLSTDKFQVFLDSTSKNGEPG 1495

QY 1493 VERSSPSKCPSLDRWMHSCSLQNRNYPSEELIKVVDVEEQLESGPHDLTETSY 1552

Db 1496 MIRSSPSQSLDLTRWVHSCPSLQDNCPSOKELTKVVSMEEQPTSEARDLMEQSY 1555

QY 1553 LPRQDLEGTPLSGSLFSDPESDPEDPARPARVGNIPSTSAKVPQPKVAESAQ 1612

Db 1556 LSRPDLEGAPLYSGSLFSDPESDPSSHRASELAVHVSMPSTSTSAKLPQFOVESAK 1615

QY 1613 SPAAHTTDTAGVNAMEESVSRKPELTASTERVNKRMSVVSGLTPPEPMLVYKARKH 1672

Db 1616 STAAVHIASHTAGVKNKSEDSVIGIEKPEVISSTGVNKRISVVASGLTPKEFMLVHKFARKH 1675

QY 1673 HITLTLITEETHVVMKTKDAEVCERTKYFLGIAGGKVVVSFYVWVTOSIKERKMLNEH 1732

Db 1676 HISLTNLISEETHVMKTKDAEVCERTKYFLGIAGGKVVVSFYVWVTOSIKERKMLDEH 1735

QY 1733 DFEVGDVNVGRNHQGPRA-----RESQDRKIFRGLGICCCYGPFTNMPTDQLEWVQLC 1787

Db 1736 DFEVGDVNVGRNHQGPRAKRESQDRKIFRGLGICCCYGPFTNMPTDQLEWVHL 1795

QY 1788 GASVWKLSEFTLGTGVHPVTVVQPDWEDNGFHAIGMCEAPVVTREWVLDSVALYQC 1847

Db 1796 GASVWKEPSLFTLSKGTHTPVVTVVQPDWEDSGFHAIGMCEAPVVTREWVLDSVALYQC 1855

QY 1848 QELDTYLIPOIPHS 1861

Db 1856 QELDTYLIPOIPRT 1869

RESULT 3

BRCL_MOUSE STANDARD; PRT: 1812 AA.

AC P48754; Q60957; Q60983;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

DE Breast cancer type 1 susceptibility protein homolog.

GN BRCL.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RN SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Embryo;

RX MEDLINE=96177659; PubMed=8634697;

RA Abel K.J., Xy J., Yin G.Y., Lyons R.H., Meisler M.H., Weber B.L.;

RT "Mouse Brcl: localization sequence analysis and identification of

RT evolutionarily conserved domains.";

RL Hum. Mol. Genet. 4:2265-2273(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6;

RX MEDLINE=96177660; PubMed=8634698;

RA Sharan S.K., Wims M., Bradley A.;

RT "Murine Brcl: sequence and significance for human missense

RT mutations.";

RL Hum. Mol. Genet. 4:2275-2278(1995).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=129/SVJ;

RX MEDLINE=96121367; PubMed=8575748;

RA Bennett L.M., Haugen-Strano A., Cochran C., Brownlee H.A.,

RA Fiedorek F.T. Jr., Wiseman R.W.;

RT "Isolation of the mouse homologue of BRCL and genetic mapping to

RT mouse chromosome 11.";

RL Genomics 29:576-581(1995).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=129/SVJ; TISSUE=Embryo;

RX MEDLINE=96067162; PubMed=7590247;

Lane T.F., Deng C., Elson A., Lyu M.S., Kozak C.A., Leder P.;

"Expression of Brcl is associated with terminal differentiation of

ectodermally and mesodermally derived tissues in mice.";

Genes Dev. 9:2712-2722(1995).

[5]

SEQUENCE OF 727-1111 FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Embryo;

RX MEDLINE=96021028; PubMed=7550308;

RA Marguis S.T., Rajan J.V., Wynshaw-Boris A., Xu J., Yin G.Y.,

RA Abel K.J., Weber B.L., Chodosh L.A.;

RT "The developmental pattern of Brcl expression implies a role in

RT differentiation of the breast and other tissues.";

RL Nat. Genet. 11:17-26(1995).

RN [6]

SEQUENCE OF 789-1250 FROM N.A.

RP STRAIN=129/SVJ;

RX MEDLINE=96163506; PubMed=8566695;

RA Schroeck E., Badger P., Larson D., Erdos M., Wynshaw-Boris A.,

RA Ried T., Brody L.;

RT "The murine homolog of the human breast and ovarian cancer

RT susceptibility gene Brcl maps to mouse chromosome 11D.";

Hum. Genet. 97:256-259(1996).

CC -!- FUNCTION: NOT KNOWN. MAY REGULATE GENE EXPRESSION. INVOLVED IN

TRANSCRIPTIONAL REGULATION OF P21 IN RESPONSE TO DNA DAMAGE (BY

SIMILARITY).

CC -!- SUBUNIT: CTIP INTERACTS SPECIFICALLY WITH THE BRCT DOMAINS (BY

SIMILARITY).

CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).

CC -!- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED IN OTIC VESICLES AT

DAY 9.5. AT DAY 10.5, THIS EXPRESSION DECREASES AND HIGH LEVELS

ARE FOUND IN THE NEUROECTODERM. AT DAYS 11-12.5, HIGH LEVELS IN

DIFFERENTIATING KERATINOCYTES AND WHISKER PAD PRIMORDIA. AT DAYS

14-17, EXPRESSION ALSO OBSERVED IN KIDNEY EPITHELIAL CELLS. IN

THE ADULT, HIGHEST LEVELS FOUND IN SPLEEN, THYMUS, LYMPH NODES,

EPITHELIAL ORGANS, AND ALVEOLAR AND DUCTAL EPITHELIAL CELLS OF

THE MAMMARY GLAND. VERY LOW LEVELS IN BRAIN, KIDNEY, AND SKIN. NO

EXPRESSION IN HEART, LIVER OR LUNG.

CC -!- DEVELOPMENTAL STAGE: IN THE MAMMARY GLAND, EXPRESSION INCREASES

DRAMATICALLY DURING PREGNANCY. LEVELS FALL DURING LACTATION AND

INCREASE AGAIN DURING POST-LACTATIONAL REGRESSION OF THE

MAMMARY GLAND.

CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.

CC -!- SIMILARITY: CONTAINS 2 BRCT DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC EMBL; U35641; AAB17113.1; -

DR EMBL; U31625; AAB17114.1; -

DR EMBL; U32446; AAA96393.1; -

DR EMBL; U36475; AAC52323.1; -

DR EMBL; U33835; AAA99742.1; -

DR MGD; MGI:104537; Brcl.

DR InterPro; IPR001357; BRCT.

DR InterPro; IPR002378; Brst_cancerI.

DR InterPro; IPR001841; Znf_ring.

DR Pfam; PF00097; zf-C3HC4; 1.

DR Pfam; PF00533; BRCT; 2.

DR PRINTS; PR00493; BRSTCANCER1.

DR SMART; SM00292; BRCT; 2.

DR SMART; SM00184; RING; 1.

DR PROSITE; PS50172; BRCT; 2.

DR PROSITE; PS00518; ZF_RING_1; 1.

DR PROSITE; PS50089; ZF_RING_2; 1.

DR Zinc-finger; DNA-binding; Nuclear protein; Anti-oncogene; Repeat.

ZN_FING 24 65 RING-TYPE.

FT DOMAIN 1585 1679 BRCT 1.

FT DOMAIN 1698 1797 BRCT 2.

```
FT VARIANT 93 93 F -> L (IN STRAIN 129/SVJ).
FT VARIANT 305 305 T -> S (IN STRAIN 129/SVJ).
FT VARIANT 319 319 P -> A (IN STRAIN 129/SVJ).
FT VARIANT 377 377 Q -> E (IN STRAIN 129/SVJ).
FT VARIANT 550 550 K -> Q (IN STRAIN 129/SVJ).
FT VARIANT 652 652 A -> P (IN STRAIN 129/SVJ).
FT VARIANT 765 765 S -> P (IN STRAIN 129/SVJ).
FT VARIANT 917 917 P -> L (IN STRAIN 129/SVJ).
FT VARIANT 933 933 C -> S (IN STRAIN 129/SVJ).
FT VARIANT 1122 1122 K -> I (IN STRAIN 129/SVJ).
FT VARIANT 1206 1206 S -> R (IN STRAIN 129/SVJ).
FT VARIANT 1212 1212 RM -> GI (IN STRAIN 129/SVJ).
FT VARIANT 1255 1255 S -> R (IN STRAIN 129/SVJ).
FT VARIANT 1261 1261 H -> N (IN STRAIN 129/SVJ).
FT VARIANT 1264 1264 V -> A (IN STRAIN 129/SVJ).
FT VARIANT 1269 1269 P -> A (IN STRAIN 129/SVJ).
FT VARIANT 1283 1283 T -> K (IN STRAIN 129/SVJ).
FT VARIANT 1337 1337 T -> N (IN STRAIN 129/SVJ).
FT VARIANT 1349 1349 P -> T (IN STRAIN 129/SVJ).
FT VARIANT 1352 1352 EG -> OR (IN STRAIN 129/SVJ).
FT VARIANT 1381 1381 S -> P (IN STRAIN 129/SVJ).
FT VARIANT 1390 1390 G -> A (IN STRAIN 129/SVJ).
FT VARIANT 1400 1400 V -> D (IN STRAIN 129/SVJ).
FT VARIANT 1503 1503 E -> Q (IN STRAIN 129/SVJ).
FT VARIANT 1549 1549 V -> A (IN STRAIN 129/SVJ).
FT VARIANT 1680 1680 T -> K (IN STRAIN 129/SVJ).
FT VARIANT 1712 1712 D -> E (IN STRAIN 129/SVJ).
FT VARIANT 1721 1721 D -> E (IN STRAIN 129/SVJ).
FT SEQUENCE 1812 AA; 198669 MW; 2291EA74150BB86A CRC64;

Query Match 49.9%; Score 4817; DB 1; Length 1812;
Best Local Similarity 55.5%; Pred. No. 2.9e-194;
Matches 1035; Conservative 261; Mismatches 498; Indels 70; Gaps 32;

QY 1 MDLSALRVEEONVINAMQKILECPICILELIKEPVSTKCDHIFKFCMLKLLNKKGPSP 60
DB 1 MDLSAVQIEOVNVLHMQKILECPICILELIKEPVSTKCDHIFKFCMLKLLNKKGPSP 60

QY 61 CPLCKNDITKRSLOBSTFVSOLVRELLIIICAFOLDTGLEVANSYNPAKKNENSPHEHLD 120
DB 61 CPLCKNEITKRSLOSTFVSQLAEBELLIIIMAAFELDTGMQLTNGFSFKRRNNSCERLINE 120

QY 121 EYVSIQSGYRNRAKRLQSPENPSLOETSLVSQSLNGLGTVRTLTKQRIQPKTSYVI 180
DB 121 EASIIQSGYRNRRRLPQVEPGNATLKD-SLIGVQLSNLGIIVRSVKKNRQTPQRKKSYYI 179

QY 181 ELGSDSSEDTYNKATYCSVGQELLOLITPQGTREISLDSAKKACFESETDVTNTEHQ 240
DB 180 ELDSDSSEETVTKPDCSVRQDELLOLQAPQAGDEGLHSAEAEACFESE-GIRNIEHHQ 238

QY 241 PSNNDLNTEKRAARHPKPYQGSVSNLHVPCGTNTNTHASSLQHNSSLLLTKDRMNV 300
DB 239 CS-DOLNPTENHATERHEKQCQTSISNVCPGCTDAHASSLQPEYSSLLLIEDRNAE 297

QY 301 KAEFCNKQKPLARSQHRNAGSKETCNDRTPTSTKKYDNLNADPLCERKEWNKQLPC 360
DB 298 KAEFCNKTKQGVIAVSQOSRWPASKGTCDNRQVPSTGEKVGPNADSLSDREKWTHPQSLC 357

QY 361 SENPROTDVPMITLNSSIOKNEWFSRDELLGSDSDSHCESNAKAVADVLVNEVD 420
DB 358 PENSQATTDVPMITLNSSVQKNEWFSRDELLGSDSDSHCESNAKAVADVLVNEVD 417

QY 421 EYSGSSEKIDLIASDPHEALICKSRVRSKSVESNIEDKIFGKYRKKASLPNLSHVTEN 480
DB 418 GGFSSSRKTDLVTPDHPHTLMCKSCRODFSKPYEDNISDKIFGKYQRKGRPHLNHYTE- 476

QY 481 LIIGAFVTEPQIIOBRPLTNKLRKRRTPSGLHPDEFIKKADLA-VQKTPMWINQGTNOT 539
DB 477 -IIGTFITEPQITQBPPTNKLKRKR--STSLQPEDFIKKADSAQVQRTPDNINQGTDL 533

QY 540 EONGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSISINMWELELN 599
DB 534 EPNEQAVSTNSCQENKTAGSNLOKESAHPTESLRKEPASTAGAKSISNSVSDLEVLN 593
```

```
QY 600 IHNSKAPKKNRLRRKSSSTRHIIHALELVVSRNLSPNCTELQIDSCSSSEIEIKKKYNQMP 659
DB 594 VHSSKAPKKNRLRRKSSIRCALPLE-PISRNPSPTCAELQIDSCSSSEIEIKKNISNOQA 652

QY 660 VRHSRNLQLMGKEPATGAKSKSNPNEQTSKRHSDTTFPELKLTNAPSGFTKCSNTSELK 719
DB 653 AGHLREPOLIEDTEPAADAKK-NEPNEHRIKRRASDAFPEEKLKMKAGLLTSCSPRKSQ 711

QY 720 EFNPSLPREEKEKLETVKVSNNAEDPKDMLSGE-RVLQTERSVESSSISLVPCTDYG 778
DB 712 GPVNPS-PORTGTEQLETRQMSDAKELGDRVLGEGPSGKTTDRSEESTSVLSVSDTYD 770

QY 779 TOESISLLEVSTLCKATEPNKVCQCAAPENKGLIHGCKSKDNRNDTEGFKYPLGHEVN 838
DB 771 TONSVSVDLDAHTVRYARTGSNOQMTQVASENPKELVHG-SNNAGSGTEGLKPLPRLHALN 829

QY 839 HSRETSIEMEESELDAQYLONTFKVSKRQSFALFSNPGNAEEECATFSAHSGSLKKQSPK 898
DB 830 LSQE-KVEMEDSELDTQYLONTFQVSKRQSFALFSKPRSPQKDC---AHSVPSKELSPK 884

QY 899 VTPECEKEENOCKNESNIKPQVFNITAGPPVVGQDKPVDNAKCSIKGSRFCLSSQF 958
DB 885 VTAKGQK-ERQGOEFEISHVQVAATAVGLVPCQEGKLAADTMC--RGCRLCPSHY 941

QY 959 RGNETGLITPNKHGLLQNPYRIPPLPIKSVKTKCKKNLLEENFEHSMSPEREMGNEN 1018
DB 942 RSGENGLSATCKSCISONSHFKQSVPIRSIKITDNKPLTEGREFERTSTTEMAVGNEN 1001

QY 1019 I-PTVSTIIRNIRENVFKEASSNINEVGSSNINEVGSSNINEGSSDENTOAELGNRG 1077
DB 1002 ILQSTVHTVSLNN-RGNACQEAGS-----GSIHEVCSTGDSFPQGLGNRG 1046

QY 1078 PKLNAMRLGVLOPEVYKQSLPGSNCKHPKIKQBEYEVQTVNTDPSLYLISDNLEQPM 1137
DB 1047 PKVNTVPLDSMQGVQCOQSVVPSD-KYLETKQEGEAVC---ADFSPLCFSDHLEQSM 1101

QY 1138 GSSHASQVCTPDDLLDDGGEIKEDTFAENDIKESSAVFSKVQKGLSRLSPSPFTTH 1197
DB 1102 -SGKVQVQVCTPDDLLDDVEKQHTSFGEGDIMERSAVFNGLSRLRESSPSPVTHAS 1160

QY 1198 LAQYRGAKKLESSENLSEDEELPCFQHLFGKVNINIPQSSTRHSTVATECLSKNTE 1257
DB 1161 KSSLHRSARAKLESSESDSDTDEDLPCFQHL--SRISNTP-ELTRCSSAVTQRMPEKAE 1218

QY 1258 ENLLSLKNSLNDCSNOVILAKASOEHLSEETKCSASLFSQCSLELDLTANTQDPL 1317
DB 1219 GTQAPWKGSSDCNNEVIMIEASQEHQFSEDPKCSGMFSQHSQSAVOGSTPNANSQDNF 1278

QY 1318 IGSSKQMRHQSESQGVGLSDKELVSDDEERGTLGEENNQERQSDSNLGEA-ASGESET 1376
DB 1279 ITPSTQRSHQCGNEEAFLSDKELISDNENMATCLEEDNDQEE--DSIIPDSEASGESET 1336

QY 1377 SVSEDCGLSQSDILITQOQRTDMQHLIKLQOEAMAELEAVLEQHGQSPNSYPSIISDS 1436
DB 1337 TLSBDC----SQSDILPTQEGATMKYNLIKLQOEAMAELEAVLEQHGQSGSHPSLLGDP 1392

QY 1437 SALEDLRNPEQSTSEKAVLTQSKSSYPISONPGLSADKFEVS-ADSSSTKNKEPGVER 1495
DB 1393 CALEDLPLVEPNMGGAAILTSKNINENPNVSONLKSACDDKPOLQHLGPTSGDDESGMR 1452

QY 1496 SSPSKCPSLDDRWYMHSCGSLQNRNYPSEBELIKVYDVDEBQOOLEESGPHDLTTSYLP 1555
DB 1453 PSPFKPLAGSRGSAHCSRHLQKRNKSPQEBELLQAPAGE---ASSEPHNSTGESCLPR 1508

QY 1556 QDEGTPYLSGSLPSD-DPESDPSEDRAPESARVINIPSTSAKLVQPKLVKAEASQSP 1614
DB 1509 RELEGTPLGSGISLFSRDPPESESPK---EPAHIGTTPASTSVLKIPQOCQAFRSAA 1564

QY 1615 AAATTTDTACYNAMESVSRKEKPELTASTERVNKRMSVYSLGTPPEFMVLYKARKHHI 1674
DB 1565 AGAD-----KAVGVGLVSKIKPELTSEERADRDIMNVVSGLTPKEVMTVQKFAEKYRL 1617
```

```
QY 1675 TLNLNITEETHVVMKTDABFCERTLKYFLGIAGGKVVYFWVTQSIKERRMLNEHDF 1734
Db 1618 TLTDATTEETHVLIKTDAEFVCERTLKYFLGIAGGKVVYFWVTQSIKERRMLNVEH 1677
QY 1735 EVRGDVVGNRHGPKKARSDQRKIFRGLIEICYGPFTNMPDQLEKVMVQLGASVYKE 1794
Db 1678 EVGDDVVVGNRHGPKKARSDQRKIFRGLIEICYGPFTNMPDQLEKVMVQLGASVYKE 1736
QY 1795 LSSFTLTGTGVHPVIVVOPDAWEDNGHFAIGQMCAPVTVREXWVLDVALYQCELDLYL 1854
Db 1737 LPSLTHTDGAHLVIVVQPSAWEDNSCPDGLQCKARLVMDVWLDVLSLSSYRCRDLDAYL 1796
QY 1855 IPQI 1858
Db 1797 VQNI 1800

RESULT 4
BRC2_MOUSE STANDARD; PRT; 3329 AA.
AC P97929; P97383; O35922;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Breast cancer type 2 susceptibility protein.
GN BRCA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=97217789; PubMed=9063750;
RA Connor E., Smith A., Wooster R., Stratton M., Dixon A., Campbell E.,
RA Tait T.M., Freeman T., Ashworth A.;
RT "Cloning, chromosomal mapping and expression pattern of the mouse
RT Brca2 gene.";
RN Hum. Mol. Genet. 6:291-300(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=97237041; PubMed=9119389;
RA Sharon S.K., Bradley A.;
RT "Murine Brca2: sequence, map position, and expression pattern.";
RN Genomics 40:234-241(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=97384941; PubMed=9242436;
RA McAllister K.A., Haugen-Strano A., Hagevik S., Brownlee H.A.,
RA Collins N.K., Futreal P.A., Bennett L.M., Wiseman R.W.;
RT "Characterization of the rat and mouse homologues of the BRCA2 breast
RT cancer susceptibility gene.";
RN Cancer Res. 57:3121-3125(1997).
RN [4]
RP SEQUENCE OF 18-200 FROM N.A.
RX MEDLINE=97075121; PubMed=8917547;
RA Rajan J.V., Wang M., Marquis S.T., Chodosh L.A.;
RT "Brca2 is coordinately regulated with Brca1 during proliferation and
RT differentiation in mammary epithelial cells.";
RN Proc. Natl. Acad. Sci. U.S.A. 93:13078-13083(1996).
RN [5]
RP SEQUENCE OF 569-625 FROM N.A.
RX MEDLINE=97341126; PubMed=9196008;
RA McAllister K.A., Ramachandran S., Haugen-Strano A., Fiedorek F.T. Jr.,
RA Wiseman R.W.;
RT "Genetic mapping of the Brca2 breast cancer susceptibility gene on
RT mouse chromosome 5.";
RN Mamm. Genome 8:540-541(1997).
CC -!- FUNCTION: MAY PARTICIPATE IN A PATHWAY ASSOCIATED WITH THE
CC ACTIVATION OF DOUBLE-STRAND BREAK REPAIR AND/OR HOMOLOGOUS
CC RECOMBINATION (BY SIMILARITY).
```

Matches 401; Conservative 292; Mismatches 742; Indels 649; Gaps 94;

Qy	72	SLOBSTRFSOLVPELKIIC-----AFOLDTGLEVANSYNFAKKNENS---PEHLKD	120
Db	466	SLEDSIAGKQMWRTSOAACLSPIRKSIFKKREPLDETGLTGVSDSMNTFTTEH	525
Qy	121	EVSTIQSMGYRRAKRLQOSEPNPS---LOETSLSVOLSNLGTVRTLRTRKQRIQOKT	176
Db	526	ACGLGILTACSQREDSICSPSVDTCGWPITLTDTSATVK--NAGLISLTKNKR-----	577
Qy	177	SVYTELGDSSDETVMKATYCVSGVDELLQIITPOQTRDEISLDSAKKAACSESTDV---	233
Db	578	-----KFIIY-SVSDASLQGRKKLQTHRQLELTNL- SAQLEASAFEVPLT	619
Qy	234	-TTEHHOPNNNDLNTTEKRAAERHEPKYQSGSSVNLHVEPCGNTTHASSILOHENSIL-	290
Db	620	FTNVNSGIPOSSD---KKRCLPNDEP-----EFLSNSFGTATSKSISIIHA	663
Qy	291	LLTKDRNMVEKAEFCNKSQOPGLARSQHNRWAGSKETC-NDRRTPTST---EKKYDLNA--	344
Db	664	LISQD-LNDKEAIVIEBKPOPYTAREADFLCLPERTCENDQKSPKVSNGKEKVLVSACL	722
Qy	345	-----DPLCERKWNKQKL-PCSENPRDTEVPW-----	372
Db	723	PSAVQJSSISFESQENPLGDHNGTSTLKLTPSKLPLSKADWVREKMKMKPEKLOCESC	782
Qy	373	---ITLNSIQTQVNWFSRDELLGDDSDHGESNAKV-----ADVLDVUNEVDEY	422
Db	783	KVNIELSKNILEVNEICILSE-----NSKTPGLLPPGENTIEVASSMKSQ	827
Qy	423	SGSEKIDLLASDPHALICKSER---VHSKSVESNIEDKI-----FGKTYRKKASLP	472
Db	828	FNQNAKI-----VIQDKQKSPFISEVAVNMNNEELFPDSSGNNAFAQVTKNCNKP	877
Qy	473	NLSHVNTLIIGAFVTEPOIIOERPL--TNKLKRRRPTSGLHPEDFIKKADLAVOKTPE	530
Db	878	DLASSVE-----LQEDLSHTQGPSLKNSPMAVDEDDVDAHAAQVLIITKDS	924
Qy	531	MINOGTNOTEQGVNMNITNSGHEKNT-KGDSIQNEKNPNPIESLEKESAFKTKAEPIS	589
Db	925	SLAVVHDYTEKSRNNTTEQHKQTEKDFKSNLSNMKSDGNSDCSKWSEP---LDPVLN	981
Qy	590	-----SISNMELNLNTHNSKAPK--NRLRRKSSTR-----HIIALELVSRNLSPP	634
Db	982	HNTGSGFRTASNKEIKLSHNKVKSKMFFKDLDEQYPTRLACIDIVNTLPLANOQKISEP	1041
Qy	635	NCTELQIDSCSSSEELKKKYNQMPVRHSRNLQLMEGREPATGAKKNKPNQOTSKRHS	694
Db	1042	HIFDLK-----SVTVYSTQSHNQSSVSH-----EDT	1067
Qy	695	DTPELKLTNAPGSFTKCSNTSELKEFVNPSLPREEKEKLETVK-----VSNNAEDP	747
Db	1068	DTAPQM-LSSKQDFHNNLTTTQKAETELSTILEESQSEFTQTRKPSHIAONTSEVP	1126
Qy	748	KDMLMSGERVLOT-ERSVRESSISLPGDTGTOESILLEVSTLGRAKTEPNKCVSOCA	806
Db	1127	GNQMV-----VLSTASKWKDTHLPLVDPDSVGQTDHSHKQFEGSA-GVKQSPFHLLDTCN	1181
Qy	807	AFENPKGLIHGCSKNRNDTE--GFKYPLIGHEVNHRSRTSIEMEESELDAQYLQNTFKVS	864
Db	1182	--KNT-----SCFLPNI NEMWFGGCSALGTKLSVNE-----A	1213
Qy	865	KROSFALFNSPGNABECCA-----TFSAHGS-----LKKOSPKVTE-----CEQK	906
Db	1214	LRAKMLFSDIENSEPSAKVPRGFSSAHDSVAVFKIKKONTEKSEFDEKSSKCOQT	1273
Qy	907	EENQ-----GKN-ESNIKPVQ-----TWNITAGFPVV---GQK	935
Db	1274	LQNNIEWTTCIFVGRNPEKYIKNTHEDSYTSSQRNNLENSDGSMSSTSG-PVYIHKGDS	1332
Qy	936	DKPVD-NAKCSIKGGSRRFCLSSQFRNGTGLITPNKHGLLQNPYRIPPLPKSFVKTKC	994
Db	1333	DLPADQSGSK-----PESQ--TOYARENTQIKENISDLTCLTEIMKAETCMKSSOKKOL	1385

[illegible]

```
RESULT 5
RBPL_PLAVB
ID RBPL1
```


Db 1875 EMMKVSAYEGMKRDHSTVSQVQDMKTIIVDELKTLNDISECSVLNVVSVIVKKVKES 1934
 QY 1284 HISEETKCSASFSCQSELEDLTANTQTPDFLIGSSKOMRHOSEQ-----GV 1333
 Db 1935 KHADYRDRANSWYESWVI-----LANYFLSDEAKISSGMFNAEMKSNFKTDLELFSV 1989
 QY 1334 GUSDKELVDDPERGTGLEENNQQSDMSNLGPAASGCESETSVSEDCSLGQSQSDILT 1393
 Db 1990 ISNSNELLKKIEQDSNDVIOKERESEQLAKDATDIYVNIKLNFEKLEAKNKEEVVS 2049
 QY 1394 TOORTMQ-----HNLTKLOEMAELEAVLBOHGSQPS----- 1426
 Db 2050 EKVRALKRLSQVEGRCHFNFHRLDNTDELENLKKWVTIYRDKKRSERESGLQEMENE 2109
 QY 1427 -NSYPS-----IISDSSALEDLRNPQST-----SEKAVLTOKSSP-----Y 1463
 Db 2110 MNTYSNITQLEGIVVSGESKEBIEKLNERNEMRNISEKISTIDSKVTMNSTIDELY 2169
 QY 1464 PISQNPGLSADKFEVSADSTSK-----NKEPGVERSPFSKC-----PSLDDRWTMHS 1512
 Db 2170 KLGKNCQAHWISLISYANTMKTSSKLLIMINK-----EKENTKCVDIKDNSSSTDGVVET 2225
 QY 1513 CSGSLQNR-NYPSQELKVVVV-----EEQLESQPHDITETSYLPRQD-----L 1558
 Db 2226 LKGFYSGKLTFTSSAIEVQONADTYSVNFAPKHEKESL--NAIRDIKKELYLFHQNSDISIV 2283
 QY 1559 EGTPLYESGILSFDPSDPSDRAPESAR-VGNIPSSTSALKVPOLKVAESAQSPAAA 1617
 Db 2284 EGG--VQNNLALY-----DKLNEEKREMDELYRNSETYKLQNE----- 2320
 QY 1618 HTDTPAG-----YNAMESVSREKPEITASTERYNKR-----SMVVSGL--TPEEPMV 1665
 Db 2321 HSTDVFKPMIELHKGMMNETNKSLEKEKLVSDNHMSMEAMIKNLKTYTPESVONI 2380
 QY 1666 -----YKARKHH-----ITLTN-----LITEETHVV 1688
 Db 2381 NNIVSVIEAEVKTLEEDRDYGDYQIVVEEHKQFSLIDRTNALMDDIIFKKNENYL 2440
 QY 1689 MKTDAEFYCERTKYFLGIAGKVV--VSFYFWTQSTIKERKMLNEHDFEVRGVDVNGRN 1745
 Db 2441 MEVNTETI-HRVNDYIEKIT-NKLVAQTEYEQILENIKQNDMLQNIPLKKYSII--EY 2496
 QY 1746 HOGPKRARES 1755
 Db 2497 FENVKKKES 2506

RESULT 6

YD86_SCHPO STANDARD: PRT: 1957 AA.
 AC Q10411;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein C1F3.06c in chromosome I.
 GN SPAC1F3.06C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,

RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Carrinquez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
 RT "The genome sequence of Schizosaccharomyces pombe."
 RL Nature 415:871-880(2002).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 270690; CAA94624.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA0617D9DA CRC64;
 Query Match 3.5%; Score 339; DB 1; Length 1957;
 Best Local Similarity 18.5%; Pred. No. 6.7e-07;
 Matches 334; Conservative 325; Mismatches 690; Indels 548; Gaps 79;
 QY 3 LSALRVEEVQNVINA-----MQKILECPI-----CLELIKEPVSTKODHIFCKFCKMLK 51
 Db 266 VSTLR--QTENSLRAECKTLQEKLEKCAINEEDSKLLEELKHNVANYSDAIVHKDKLIED 323
 QY 52 LNK-----KGPSQCLPKNDITKR-----SLOES--TRFSOLVVELLKIICAFOLD 96
 Db 324 LSTRISFDNLKSERDPLSLKNEKLEKLLRNTIGSLKDSRTSLSOLEEEMVEL----- 376
 QY 97 TGLEVANSYNFAKKENNSPEH--LKDEVSIIQSVMYGNRAKRLQLQSEPNPNSLOETSLV 154
 Db 377 -----KESNRTIHSQTLDAESKLSFQEENKSLKSGIDEXQNLSKDKVVK 423
 QY 155 QLSN-----LGTVRTLTQRIOQPKTSVYIELGSD-----SSEDVTNKAT 195
 Db 424 QVSSQLEEARSSLAHATGKLAENINSERDFQNKIKDFEIQDLRACLUNSSSNELKESKA 483
 QY 196 YCSVGDOELLQITPOGTRDEISLSAKKAACEFSETDVTNTEHHQPSNNDLNTTEKRAE 255
 Db 484 LIDKQOELNNLRQ--IKEQKKVSESTQSSLSQSLQORDILN-----EKK--- 525
 QY 256 RHPKEYOGSSVSNLHVPCGTNTTHASSLQHENSSLLTKDRMNVKAEFCNKSQOPGLAR 315
 Db 526 KH-EVYE--SQLNELKE--LQTEISNSELHSSOL----- 555
 QY 316 SQHNRWAGSKETCNDRRTPSTKXKVDLNADPLCE--RKNNKQKLPCESENPRDTEVPWI 373
 Db 556 ---STLAAEKEAAVATNNELSEKNSLQT--LCNAPQKLAESVMQLKENQN-----FS 605
 QY 374 TLNASSQKVNWFERSRDELGSDSHGSESNKAKVADVLDVLN-----EVDEYSGSSE 427
 Db 606 SLDTSFKKLNE---SHOEL---ENNHOTITKQLKDTSSKLOQLQOLERANFEQKSTLSDE 659
 QY 428 KIDLLA-----SDPHEALICKSERVSHSKSVESNIEDIFGKTYRK-----KASLPNL 474
 Db 660 NNDLRYTKLLKLEESKSLIKQEDV--DSLEKNIQ--TLKEDLRKSEALRFSKLEAKNL 715
 QY 475 SHVTENLIIGAFTVEPTQIIQERPLTNKLRKRRTPSGLHPEDFTFKADLAVQKTP--EMI 532

```
Db 716 REVINDL----KGKHETLEAQRNDLHSLSDAKN-TNAILSELTKSSBEDVKRLTANVETL 771
QY 533 NQGTNOTEQN-----GOVMNITNSGHENKTKGDSIQNEKNPNPTLESLEKSAFKTKAEPI 587
Db 772 TQDSKAMKQSFSLVNSQYSISNLYHELRODHVNMQSONN-----TLLESSEKLTDCENL 827
QY 588 SSS-----ISNWELELNTHNSKAPKKNRLRRKSSRPHTHALELNVYSRNLS----- 632
Db 828 TQQNMTLIDNVQKLHMKHVNQESKVSLEKEVNGKLSLDLKNLRSLSNVAISDNDQIILTQL 887
QY 633 ---PPNCTELQIDSCSSSEETKKKKYNOMPVRHRSRNQL-----MEGKEPATCAKSKNK 683
Db 888 AELSKNTVDSLEQESAQLNSGLSLAEKQ-LLHTENELHRLDKLTGKLIKIESKSDSL 946
QY 684 PNEOTSRHRSDTPELKLTAAPGFTKCSNTSELKFEVNPSPLPREEKEE-----KLSTVK 739
Db 947 GKLLTAQEE---ISNLKEENMSQAITSVKSKLDETLSKSSKLEADIEHLKKNVSEVE 1003
QY 740 VSNNAEDPKMLMSGRVLOTERSVSSSISLVPGTDTGQTESLSLLEVTLGKAKTEPN 799
Db 1004 VERNA-----LLASNERLMD----- 1018
QY 800 KCVSOCAAFENPKGLIHGCSKDNNDTEGPKYPLGHEVNHRSRETSIEMESSELDQAQYLQN 859
Db 1019 -----DUKNNGENTA-----SLQTEIEKKRAENDD--LQS 1046
QY 860 TKFY--SKROSFAFNSPNAGAEBCATFSAHSGSLKKQSPKVTPECKQKBEENOGK--NES 915
Db 1047 KLSVVSSEYENALLISSQTN-----KSLDKTNQLKYIEKNVQKLLDEK 1090
QY 916 NIKPVQTVNITAGPPVVGQDKDPVDNAKCSIKGGRFCLSSQFRGNETGTITPNKHGLLQ 975
Db 1091 DORNVELEELTSKYGLKEENEAQIKDELLALRKSK-----KQHDLCAL 1133
QY 976 NPYRIPPLFPIKSPVTKCKKNLLEENFEHMSPEREMGNENIPSTVTSIRNNIRENV 1035
Db 1134 N-----FV-----DOLKEKSDALE-QLTNEKNELLVSLSEQSSNNEAL 1170
QY 1036 FKEAS--SSNINEVGSSTNEVGSINEIGSSDENIQALGRNGRGPKNLAMLRLGLVQPEV 1093
Db 1171 VEERSDLANLSDMKKSLSDSNVISRSDLVVRNDEL-----DT 1211
QY 1094 YKQSLPSGCKHPEIKQEEVVEVQTV---NTDFSPYLIIDNLEQPMGSSHASQVCSETP 1150
Db 1212 LKDKDLSLSTQYSEV-CQDRDDLLSLDKGCEESFNKYAVS-----LRELCTKSE 1259
QY 1151 DD-----LLDDGEIKEDTSPA-----ENDIKESSAV-PSKSVQKGELSPSPFT 1194
Db 1260 IDVPVSEILDONFVFNAGNFSELSRLTVLSLENYLDFAFNQVNFKKMELDNRLFTTDAEFT 1319
QY 1195 -----HTHLAQGYRRG--AKKLESSEENLSSEDEELPCFQHLFGKVNPNIPSQST 1242
Db 1320 KVVADEKLEQHEDDHLIQRGDEKALKDSEKNFLRKEAEM-----TENIHSLEE 1369
QY 1243 RHTVATEC--LKNTEENLLS---LKNSLDNCNQVILAKASQEHHLSETKCSASF 1297
Db 1370 GKBEETKEIAELSRLEDNQLATNKLKNQD-----HLNQEIIRLKDVLK 1414
QY 1298 SOGS---ELEDLTANTTQDPFLIGSSKOMRHOSESOGVGLSDKELSDDEERGCTGLEEN 1354
Db 1415 EKESLIISLESLSNQKQESSLLDAKNELEH-----MLDDTSRKNSSLMEX 1461
QY 1355 NQE-EQSMDSNLGAEAGCESETSVSDECSGLSSQSDILATQORDTMQHNLIKLOQEMAE 1413
Db 1462 IESINSLDDKSFELASAVEKLGALQK-----LHSESLSLMENIKSQLQEAKEIQVDEST 1517
QY 1414 LEAVEBHGQSPNSY-----PSIISD--SSALEDLRN--PGQSTSEKAVLTSQKSEY 1463
Db 1518 IQE-LDHEITFASKNNYBEGKLNKSDSIIRLSENIEQNLNLLAEKSAVKRLSTEKESEIL 1576
QY 1464 PISONPEGLSADKFEVSADSSSTSKNKEPGVGRSPSKCPSLDORWYMHSCSGSGLONRNP 1523
Db 1577 QFNSRLADLEYHKHQSVESELSGRSKLK-----LASTTEELQLAENE 1616
```

RESULT 7

MLP1_YEAST

ID MLP1_YEAST STANDARD; PRT; 1875 AA.

AC Q02455;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE MYOSIN-like protein MLP1.

GN MLP1 OR YKR095W OR YKR415.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C;

RX MEDLINE=93247549; PubMed=8483450;

RT Koelling R., Nguyen T., Chen E.Y., Botstein D.;

RA "A new yeast gene with a myosin-like heptad repeat structure.";

RL Mol. Gen. Genet. 237:359-369(1993).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=94205265; PubMed=8154186;

RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,

RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;

RT "The complete sequence of a 15,820 bp segment of Saccharomyces

RT cerevisiae chromosome XI contains the UBI2 and MLP1 genes and three

RT new open reading frames.";

RL Yeast 9:1349-1354(1993).

CC -!- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA

CC REPAIR.

CC -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.

CC -!- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MLP1".

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; L01992; AAA34783.1; -;

CC DR EMBL; X73541; CAA51948.1; -;

CC DR EMBL; Z28320; CAA82174.1; -;

CC DR PIR; S38173; S38173.

CC DR SGD; S0001803; MLP1.

CC Coiled coil; DNA repair.

KW DOMAIN 69 487

FT COILED COIL (POTENTIAL).

FT FT DOMAIN 531 1678

FT COILED COIL (POTENTIAL).

FT FT DOMAIN 1834 1866

FT CONFLICT 301 301

FT R -> A (IN REF. 1).

SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C906867 CRC64;

Query Match

3.4%; Score 327; DB 1; Length 1875;

Best Local Similarity 18.7%; Pred. NO. 2e-06;
Matches 300; Conservative 269; Mismatches 603; Indels 432; Gaps 70;

Qy	4	SALRVEEVQNVINAM-----QKILECPICILELKPVSTKCDHIKFCFCKMLKLNQK	55
Db	438	AALLLEHTSNEKNAKVELNAKNOKEVENDLQTLTKORLDLCRQIQYLLITNSVSNDS	497
Qy	56	KGP-----SQCPLCKNDITKRSIQESTFSOLVE-----ELLKII--CA	92
Db	498	KGPURKEEIOFIQINMOEDOSTITESDSOKVYTERLVFKNIIQLOEKNAELLKVVNRLA	557
Qy	93	FOLDGTLEYANSYFAKKNENSPHEHLKDVSTIQS--MGYRNRAKRLKQSEPNPSLQET	150
Db	558	DKLES-KEKSKOSQLOKIESETVNEAKEAIITLKSEKMDLESIRIE-LOKELBELKTSVP	615
Qy	151	SLSVOLSNLGTVTLTKTOKIOPO-----KTSV	178
Db	616	NEDASYGNVTIKQLTETKRDLESOVQDLQTRISQITRESTENMSLLNKEIQDLYDSKDI	675
Qy	179	YIELGSSS-----EDTVNKA	194
Db	676	SIKLGKSKSRIILAEBRFKLLSNYLDLTKAENDQURKRPDYLQNTILKQDSKTHETINEY	735
Qy	195	TYC-----SVGQDQLLOITPO-----GTRDEISLDSAKKAACEFSETDVTVTEHHQP	241
Db	736	VSCKSKLSIVETELLNKEEQKLRVHLEKNLQELNKLSPKQSLRIM--VTQLOTLQK	792
Qy	242	SNNDLNTTEKRAAEHPKPYQSSVSNLHVPCGNTHTASSLOHENSLLLTTKDRMVE-	300
Db	793	EREDLLEETKRSQCKKIDELE-DALSEKKETQSQRDHIKQILEEDNS-----NIEW	843
Qy	301	--KAEPCKN-----SKQPLGARSOHNRWAGSKETCNDRTPTSEKKVD--LNAD	345
Db	844	YQNKIEALKKDYESVITSVDSQOTDIEKLQYKVKLEKEIEEDKIRLHTYNVMDDETINDD	903
Qy	346	PLCKERKWNKQKLPCE--NPRDTEJDPWITLNSIOKVNEMTSRSDLELG-----	394
Db	904	SL--RKELEKSKINLTDAYSQIKKEYKOL-YETTSQLOOTN--SKLDESFKDTNOIKN	957
Qy	395	-SDSDHGDGESANAKVADVLVDINEVD-----EYGSSEKIDLLASDPHEALICKSE	445
Db	958	LTDKTSLSLEDKISLLKEQMFNLNLDLQKKGMWEKADFKKRISILQNNKNEVEAVKSE	1017
Qy	446	-----RVHKSVESNIEDKIFGKTVYRKASI-PNLSHVTENLIIGAFVTE	489
Db	1018	YESKLSKIQNDLQOITYANTQNNYEOEL-----QKRADEVSKTISELQOLHTYKGOVK	1072
Qy	490	PQIIQERLPTNKLKRKRPTSGLHPDEFITKADLAVQKTPEMINQGTNOTEQNGVMMIT	549
Db	1073	TLNLSRDQLENALKENEKSWSS-QKESLLEQLDLSNRIEDLSQNKLLYDQI-QIYTAA	1130
Qy	550	NSGHNKTKGDSIQN-----EKNPNPISLEKESAFKTKAEPISGSSISNMELEL----	599
Db	1131	DKEVNSTNGPGLNLTILTLRRERDILT--KVTVAERDAKMLRQKISLMDVLEQDARTK	1188
Qy	600	IHNSKAPKNR-----LRRKSSRTHALVELVSRNLSPPCNCTELQIDSCSSSE-----	652
Db	1189	LDMRSVEKENHSSIIQOHDDEIKUNQJNLRESNITURN--ELENNNKKKELQSELDK	1246
Qy	653	KKYNQMPVRHSRNLQIMECKEPATGAKSKNKPNEQTSKRHDSFDTPELKLITNAPGSETK	712
Db	1247	LQONVAPI-----ESELTKALYSQMEKEQ-----ELKLAKE	1276
Qy	713	SNITSLEKFPVNSLPREEKEKLET---VKVYSNNAEDPKQMLMSGERV-----	757
Db	1277	---EYHRWKKRSQDILEKHEQLSSDYKEKLESETENILELENKERNQGAEBEKFNRLR	1333
Qy	758	LOTERSVESSISLVPDGTGYQTESISLLEYSTLTKGAKTEPNKVCVSOCA-----FENPKG	813
Db	1334	RQAKERLTKSKUS-----QDSLT-EQVNSURDAKWLNSLSANARIELOQNAK-	1382
Qy	814	LIHGCSKDRNRNDEGPKYPLGHEVHNSRETSTEMESELDAQYLQN-----TFK--VS	864

Db 1383 - - - - - VAQGNQQLLEA IRKIQEDAEKASRLQAKLBEESTTSVESTINGUNEETITLKEEIE 1437
 Qy 865 KRGSF--ALFSPNPGNAEECATFSAHSGSLKK--OSPVTPECEQKEQNKESNTKPV 920
 Db 1438 KQRIQOQLOATSANQNDLSNIVE---SMKKSPEEDIKFP---IKETQEVNEKILEAQ 1491
 Qy 921 QTVNITAGFPVVGOKDPVDNAKCSIKGSRFLCSQSPRGNETGLITPNKHGLLQNPYI 980
 Db 1492 ERLNOPSINMEEIKKKWESEHEQEV-----SQKIREAEEA-----LKKRIRL 1534
 Qy 981 PPLFPPTKSFVKYCKKNLLLEENPEEHSMPEREMENINIPSTVTSISNNIR----- 1032
 Db 1535 PTEEKINKIIR--KKDELEKEFEE---KVEERIKSMQSGEIDVYLRKQLAEKVQEKQK 1589
 Qy 1033 --ENVFKEASSNNINVGSGSTN-----EYVGSSINETGSSDENIQAEIGNRGPK 1079
 Db 1590 ELENEYNKKLQELKDVPISSHISDDERDKLRAETESRUREEFNNELQAIKKKSFDEG-K 1648
 Qy 1080 LNAMLRLGVLOPEVYKOSLPGSCNCKHPETKQYEEVVQTVNTQFSPYLIISDNLEQP-MG 1138
 Db 1649 QAAMKMTLLERKLAKWESQLS-----TKQSAESPSPKSYN-----NVQNPLIG 1692
 Qy 1139 SSNASQVCSETPDDLDDGEIKEDTSFAENDIKESSAVFSKSVQKGELSRSPSPFTHTHL 1198
 Db 1693 LPRKIEENSPFPNPLISGE-----KLLKLNKSSSGGNPFT-----SPSPNKH--- 1737
 Qy 1199 AQGYRGAKKLESSEENLSEDEELPCQHLFLFGKVNNIPOS---TRHSTVATECLSKN 1255
 Db 1738 ---LQNDNOKRESLANKTDPPTHLPSF-----NIPASRLISSSTLSLTD---T 1781
 Qy 1256 TEENLLSLKNSLNCDSNQVILAKAQE-----HHLSEETKCSASFSSQCSSEL- 1303
 Db 1782 NDELSNPNPAQKSSNRNVQSEDETEKKKEPVKRGAEATEEQTKSNKRPI-DEVGELK 1840
 Qy 1304 ---EDITANTNTQDPFLIGSSKOMRHQSESQGVGLSDRELVSDD 1344
 Db 1841 NDEDDTTENIN-----ESKKIKTEDEEE-----KETDKVND 1872

 RESULT 8
 BRC2_HUMAN STANDARD; PRT; 3418 AA.
 AC P51587; Q13879; O00183; O15008;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 18-JUN-2002 (Rel. 41, Last annotation update)
 DE Breast cancer type 2 susceptibility protein.
 GN BRCA2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 [1]
 RN BRCA2.
 RP MEDLINE=96112016; PubMed=8524414;
 RX Wooster R., Bignell G., Lancaster J., Swift S., Seal S., Mangion J.,
 RA Collins N., Gregory S., Gumbs C., Micklem G., Barfoot R., Hamoudi R.,
 Patel S., Rice C., Biggs P., Hashim Y., Smith A., Connor F.,
 RA Arason A., Gudmundsson J., Fienen C., Kelsell D., Ford D., Tonin P.,
 RA Bishop D.T., Spurr N.K., Ponder B.A.J., Eeles R., Peto J., Devilee P.,
 RA Cornelisse C., Lynch H., Narod S., Lenoir G., Egilsson V.,
 RA Barkworth R.B., Easton D.F., Bentley D.R., Futreal P.A.,
 RA Ashworth A., Stratton M.R.;
 RT Identification of the breast cancer susceptibility gene BRCA2.";
 RL Nature 378:789-792(1995).
 [2]
 RN BRCA2.
 RP MEDLINE=96172838; PubMed=8589730;
 RX Tavtigian S.V., Simard J., Rommens J., Couch F., Shattuck-Eidens D.,
 RA Neuhausen S., Merajver S., Thorlacius S., Offit K., Stoppa-Lyonnet D.,
 RA Balenger C., Bell R., Berry S., Bogden R., Chen Q., Davis T.,
 RA Dumont M., Frye C., Hattier T., Jammulapati S., Janecki T., Jiang P.,
 RA Kehrer R., Leblanc J.-F., Mitchell J.T., McArthur-Morrison J.,
 RA

RA Nguyen K., Peng Y., Samson C., Schroeder M., Snyder S.C., Steele L.,
 RA Stringfellow M., Stroup C., Swedlund B., Swensen J., Teng D.,
 RA Thomas A., Tran T., Tran T., Tranchant M., Weaver-Feldhaus J.,
 RA Wong A.K.C., Shizuya H., Eyfjord J.E., Cannon-Albright L., Labrie F.,
 RA Skolnick M.H., Weber B., Kamb A., Goldar D.E.:
 RT "The complete BRCA2 gene and mutations in chromosome 13q-linked
 RT kindreds.";
 RL Nat. Genet. 12:333-337(1996).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP Hunt S., McMurray A., Williamson H.;
 RA Submitted (OCR-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP VARIANT OC HIS-2787, AND VARIANTS ASN-372; MET-1915 AND VAL-2466.
 RX MEDLINE-96275740; PubMed=8665505;
 RA Takahashi H., Chiu H.-C., Bandera C.A., Behbakht K., Liu P.C.,
 RA Couch F.J., Weber B.L., Livolsi V.A., Furusato M., Rebane B.A.,
 RA Cardonick A., Benjamin I., Morgan M.A., King S.A., Mikuta J.J.,
 RA Rubin S.C., Boyd J.;
 RT "Mutations of the BRCA2 gene in ovarian carcinomas.";
 RL Cancer Res. 56:2738-2741(1996).
 RN [5]
 RP VARIANTS ASN-372; ASP-991; SER-1147; MET-1915 AND CYS-2034.
 RX MEDLINE-96241588; PubMed=8673091;
 RA Couch F.J., Farid L.M., Deshano M.L., Tavtigian S.V., Calzone K.,
 RA Campeau L., Peng Y., Bogden B., Chen Q., Neuhausen S.,
 RA Shattuck-Eidens D., Godwin A.K., Daly M., Radford D.M., Sedlacek S.,
 RA Rommens J., Simard J., Garber J., Metrajer S., Weber B.L.;
 RT "BRCA2 germline mutations in male breast cancer cases and breast
 RT cancer families.";
 RL Nat. Genet. 13:123-125(1996).
 RN [6]
 RP VARIANT GLO-3095.
 RX MEDLINE-96225456; PubMed=8640235;
 RA Lancaster J.M., Wooster R., Mangion J., Phelan C.M., Cochran C.,
 RA Gumbs C., Seal S., Barfoot R., Collins N., Bignell G., Patel S.,
 RA Hamoudi R., Larsson C., Wiseman R.W., Berchuck A., Iglehart J.D.,
 RA Marks J.R., Ashworth A., Stratton M.R., Futreal P.A.;
 RT "BRCA2 mutations in primary breast and ovarian cancers.";
 RL Nat. Genet. 13:238-240(1996).
 RN [7]
 RP VARIANTS
 RX MEDLINE-96225457; PubMed=8640236;
 RA Teng D.H.-F., Bogden B., Mitchell J., Baumgard M., Bell R., Berry S.,
 RA Davits T., Ha P.C., Kehrner R., Jammulapati S., Chen Q., Offit K.,
 RA Skolnick M.H., Tavtigian S.V., Jhanwar S., Swedlund B., Wong A.K.C.,
 RA Kamb A.;
 RT "Low incidence of BRCA2 mutations in breast carcinoma and other
 RT cancers.";
 RL Nat. Genet. 13:241-244(1996).
 RN [8]
 RP VARIANT ASN-2415.
 RX MEDLINE-96225458; PubMed=8640237;
 RA Miki Y., Katagiri T., Kasumi F., Yoshimoto T., Nakamura Y.;
 RT "Mutation analysis in the BRCA2 gene in primary breast cancers.";
 RL Nat. Genet. 13:245-247(1996).
 RN [9]
 RP VARIANT BC ASP-2089, AND VARIANT VAL-3412.
 RX MEDLINE-97294396; PubMed=9150152;
 RA Vehmanen P., Friedman L.S., Eerola H., Sarantausta L., Pyrhonen S.,
 RA Ponder B.A.J., Mhonen T., Nevanlinna H.;
 RT "A low proportion of BRCA2 mutations in Finnish breast cancer
 RT families.";
 RL Am. J. Hum. Genet. 60:1050-1058(1997).
 RN [10]
 RP VARIANT BC AND PANCREAS CANCER TRP-554.
 RX MEDLINE-98316775; PubMed=9654203;
 RA Ganguly T., Dhulipala R., Godmellow L., Ganguly A.;
 RT "High throughput fluorescence-based conformation-sensitive gel
 RT electrophoresis (F-CSEGE) identifies six unique BRCA2 mutations and an
 RT overall low incidence of BRCA2 mutations in high-risk BRCA1-negative
 RT breast cancer families.";
 RL Hum. Genet. 102:549-556(1998).

RN [11]
 RP VARIANTS BC L-32; R-53; L-81; R-201; A-211; S-222 AND T-3118.
 RX MEDLINE-98272917; PubMed=9609997;
 RA Katagiri T., Kasumi F., Yoshimoto M., Nomizu T., Asaishi K., Abe R.,
 RA Tsuchiya A., Sugano M., Takai S., Yoneda M., Fukutomi T., Nanba K.,
 RA Makita M., Okazaki H., Hirata K., Okazaki M., Furutsuma Y.,
 RA Morishita Y., Iino Y., Karino T., Ayabe H., Hara S., Kajiwara T.,
 RA Houga S., Shimizu T., Toda M., Yamazaki Y., Uchida T., Kunitomo K.,
 RA Sonoo H., Kurebayashi J.-I., Shimotsuna K., Nakamura Y., Miki Y.;
 RT "High proportion of missense mutations of the BRCA1 and BRCA2 genes in
 RT Japanese breast cancer families.";
 RL J. Hum. Genet. 43:42-48(1998).
 RN [12]
 RP VARIANTS OC PRO-75; HIS-2502 AND HIS-3098.
 RX MEDLINE-96455732; PubMed=10486320;
 RA Gayther S.A., Russell P., Harrington P., Antoniou A.C., Easton D.F.,
 RA Ponder B.A.J.;
 RT "The contribution of germline BRCA1 and BRCA2 mutations to familial
 RT ovarian cancer: no evidence for other ovarian cancer-susceptibility
 RT genes.";
 RL Am. J. Hum. Genet. 65:1021-1029(1999).
 RN [13]
 RP VARIANTS HIS-289; ASN-372; ASP-991 AND VAL-3412.
 RX MEDLINE-99254821; PubMed=10323242;
 RA Li S.-L., Tseng H.-M., Yang T.-P., Liu C.-H., Teng S.-J.,
 RA Huang H.-W., Chen L.-M., Kao H.-W., Chen J.-H., Tseng J.-N., Chen A.,
 RA Hou M.-F., Huang T.-J., Chang H.-T., Mok K.-T., Tsai J.-H.;
 RT "Molecular characterization of germline mutations in the BRCA1 and
 RT BRCA2 genes from breast cancer families in Taiwan.";
 RL Hum. Genet. 104:201-204(1999).
 RN [14]
 RP VARIANTS BC, AND VARIANTS
 RX MEDLINE-99138688; PubMed=9971877;
 RA Wagner T.M.U., Hirtlenhner K., Shen P., Moeslinger R., Muhr D.,
 RA Fleischmann E., Concin H., Doeller W., Heid A., Lang A.H., Mayer P.,
 RA Petru E., Ropp E., Langbauer G., Kubista E., Scheiner O.,
 RA Underhill P., Mountain J., Stierer M., Zielinski C., Oefner P.;
 RT "Global sequence diversity of BRCA2: analysis of 71 breast cancer
 RT families and 95 control individuals of worldwide populations.";
 RL Hum. Mol. Genet. 8:413-423(1999).
 CC -!- FUNCTION: MAY PARTICIPATE IN A PATHWAY ASSOCIATED WITH THE
 CC ACTIVATION OF DOUBLE-STRAND BREAK REPAIR AND/OR HOMOLOGOUS
 CC RECOMBINATION.
 CC -!- SUBUNIT: INTERACTS WITH RAD51.
 CC -!- TISSUE SPECIFICITY: HIGHEST LEVELS OF EXPRESSION IN BREAST AND
 CC THYMUS, WITH SLIGHTLY LOWER LEVELS IN LUNG, OVARY, AND SPLEEN.
 CC -!- DISEASE: BREAST CANCER (BC) IS AN EXTREMELY COMMON MALIGNANCY,
 CC AFFECTING ONE IN EIGHT WOMEN DURING THEIR LIFETIME. A POSITIVE
 CC FAMILY HISTORY HAS BEEN IDENTIFIED AS MAJOR CONTRIBUTOR TO RISK OF
 CC DEVELOPMENT OF THE DISEASE, AND THIS LINK IS STRIKING FOR EARLY-
 CC ONSET BREAST CANCER. MUTATIONS IN BRCA2 ARE THOUGHT TO BE
 CC RESPONSIBLE FOR SOME INHERITED BREAST CANCER. IT IS LINKED WITH
 CC MALE BREAST CANCER.
 CC -!- SIMILARITY: CONTAINS 8 BRCA2 REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X95152; CAA64484.1;
 CC EMBL; X95153; CAA64484.1; JOINED.
 CC EMBL; X95154; CAA64484.1; JOINED.
 CC EMBL; X95155; CAA64484.1; JOINED.
 CC EMBL; X95156; CAA64484.1; JOINED.
 CC EMBL; X95157; CAA64484.1; JOINED.
 CC EMBL; X95158; CAA64484.1; JOINED.
 CC EMBL; X95159; CAA64484.1; JOINED.
 CC EMBL; X95160; CAA64484.1; JOINED.
 CC EMBL; X95161; CAA64484.1; JOINED.

DR EMBL; X95162; CAA64484.1; JOINED.
DR EMBL; X95163; CAA64484.1; JOINED.
DR EMBL; X95164; CAA64484.1; JOINED.
DR EMBL; X95165; CAA64484.1; JOINED.
DR EMBL; X95166; CAA64484.1; JOINED.
DR EMBL; X95167; CAA64484.1; JOINED.
DR EMBL; X95168; CAA64484.1; JOINED.
DR EMBL; X95169; CAA64484.1; JOINED.
DR EMBL; X95170; CAA64484.1; JOINED.
DR EMBL; X95171; CAA64484.1; JOINED.
DR EMBL; X95172; CAA64484.1; JOINED.
DR EMBL; X95173; CAA64484.1; JOINED.
DR EMBL; X95174; CAA64484.1; JOINED.
DR EMBL; X95175; CAA64484.1; JOINED.
DR EMBL; X95176; CAA64484.1; JOINED.
DR EMBL; X95177; CAA64484.1; JOINED.
DR EMBL; U43746; AAB07223.1; -.
DR EMBL; Z74739; CAA98995.2; -.
DR EMBL; Z73359; CAA97728.1; -.
DR Genew; HGNC:1101; BRCA2.
DR MIM; 600185; -.
DR InterPro; IPR002093; BRCA2_repeat.

Query Match 3.4%; Score 327; DB 1; Length 3418;
Best Local Similarity 18.1%; Pred. No. 4.4e-06;
Matches 427; Conservative 330; Mismatches 765; Indels 838; Gaps 107;

QY 2 DLSALRVEEQVNIWAKIL-----ECPI-----CLEIKPEVSTKCDH----- 41
Db 97 ELDFKFLDLGRVNPNSRHSKLRVTKMKDQADVSCPLLNSCLS--ESPVVLOCTHTVTPQ 154
QY 42 ----IFC-----KF-----CMKLKLNQ 54
Db 155 RDSVVGCSLFHTPFVKVGRQTPKHISBSLGAEDPDMWSWSSSLATPTLSSTVLIVRNE 214
QY 55 KGPSCQCLPKNDIT-----KRSLOSTRFSQVLELLKIIICAFQDLDGTLEYANS 104
Db 215 EASETVFP---HDTTANVKSYSFNHDESLKKNDRFIASVTD-----SENTNOREAAS 263
QY 105 YNFAKKENNS-----PEHLKDEVSIIQSNGYRNRKAKRLLQSEFPNLSQET 150
Db 264 HGFGTSGNSFKVNSCKDHIGKSPNVLEDEY-----YETVVDTSDEED 306
QY 151 SLSVOLDS-----NLGTVRLTRTKRIQPOKTSVYIELGSDSDSEDTVN--KATYCSVGDOE 203
Db 307 SFLCSFKCRKNLQKVTSKTKKI-----PHEANADECEKSKQVKREKYSFVSEVE 359
QY 204 LLOITP-----OQTRDEISLDSAKKAACEF 228
Db 360 PNDTDPDLSNVAHQPFESGSGDKISKEVVPSLACIEWSOLTLGSLNGAQMEKIPLLHISSC 419
QY 229 ----SETDVTNTEHHQPSN-----NDLNTTEKRAARHPKRYQ 263
Db 420 QDNISEKDLDTENKRRKDDFITSENSLPRISSLPKSEKPLNEETVNVNRKDEEQHLESHT- 478
QY 264 SSVSNLHVPCQTHASSLOHENSLL-LFKDRMNVKEAEFCNKSQKOPGLARSQHRWA 322
Db 479 DCILAVQAISGTSVPASSFOGIKKSIPRIRESPKETFNASFSGHMTDPNFKKTEASES 538
QY 323 GSK--ETCNDRR-----TPSTEKKVDLNDAPLCE--RKENWKN----- 355
Db 539 GLEIHTVCSQKEDSLCPNLIDNGSWPATTQNSVALKNAGLISLTKKTKNFYIAIHDET 598
QY 356 ----QKLP-----CSENPRDTEVP 371
Db 599 SVYGGKKIPKQKSELINGSAQFEANAFEAPLTFANADSGLLHSSVKRCSQSN--DSEEP 655
QY 372 WITLASSITQKNEWFSRSD-----ELLCSD-----DS-----HDG 401
Db 656 TISLTSSTFCTILKCSRNETCSNTNVTISQDLDYKAEKCNKEKQLQFITPPADSLSCLOEG 715
QY 402 EESNAKVADVLDLVNEY-----DEYSGS-----SEKIDLLASDPHEALICKSER 446
Db 1632 SIFLKVKHVENVEKETAKSPATCYTNQSPYSVIENSALAFYTCGRSKTSVSQTSLSLEAKK 1691

Db 716 QCENDPKSKVSDIKKEEVLAALAAHPVQHSKVEYSDTFQSQSKSLLYDHENASTLILPT- 774
QY 447 VHSKSVESNI-----EDKIFGKTYRKXKASLNLNLSHVTENLIGAF----- 486
Db 775 --SKDVLNLMVIRGKESYKMSDKLKGNNYESDVLTGKTNIPMEKNQDVCALNENYKNVE 832
QY 487 -----VTEP-----QIIQERPLTNLKKRRPTS-----GLHPE-----DFT- 518
Db 833 LLPEKYMRVASPSRKVQFNQNTNLRVQKNOEETTSISKITVNPDSBELSDNENFVF 892
QY 519 ----KKADLAVQKTP-----MINQGTQTEQNGQVMNTNS----- 551
Db 893 QVANERNNALGNTKELHETDLTCVNEPIFKNSTMLVYDGTGDKQATQV-SIKKDLVYVL 951
QY 552 GHEHK-----TKGDSIONENKPNPIESLEKESAFKTK-----AEPIS-----S 590
Db 952 AEENKNSVKQHKMTLQDLKSDISLNDIKPEKNNDYMNKNWAGLGLPISNHSFGGSFRT 1011
QY 591 ISNMELELNIIHNSKAPKKNLRK-----SSTRHIALELVVSRNLSPPNC----- 636
Db 1012 ASNKEIKLSEHNI---KSKMFFKDIIEQYPTSLACVEIVNTLALDNQKKLSKPOSINTV 1068
QY 637 -----TELQIDSCSSSEIEKKKYNQMPVRRHNL---QLMEGKEPATGAKSKSNRPEQT 688
Db 1069 SAHLQSSVWVSDCKNSHITPQMLFSKQDFNSHNHLTPSQKAEITELSTILESGSQFET 1128
QY 689 SKRDS-----DTP--PELKLTNAPGFTKCSNTSELKEFVN-PSLPREEKEEKLTVKV 740
Db 1129 QFRKPSYILOKSTFEVPEQNOMTKTITSECRD-ADLHVIMNAPSIGQVDSKOPE---- 1183
QY 741 SNNAEDPKDMLSGERVQLQTERSVESSSISLVPGTQYQTESISLLEVSTLGCKAKTEPNK 800
Db 1184 -----GTVE-----IKRFAGLLKNDCK 1202
QY 801 CVSOCAAFENPKGLIHGCKDNMRNDTEGFKYPLGHEVNHRSRETSTIEMESELDAQYLONT 860
Db 1203 SASGLYTDENEVGF-----RGFYSAGHTKLVNSTEA-----LQKA 1237
QY 861 FKYSKRQSFALFSPGNAEEBECATFSAHSGSLKK-----QSPKVTPECE---QKEENQGN 913
Db 1238 VK-----LFSDIENISEETSA-EVHPISLSSKCHDSVVMFKIENHNDKTVSEKNN 1288
QY 914 ESNIKPQVTNITAGF---PVWGQKDPVDNAKCSIKGSRFCLSSQFRG----- 960
Db 1289 KCQLLONNIEMTGTVEEITENYKRNTEDEKNKYTAASRNSHNLFEFGSDSSKNDTVC 1348
QY 961 ---NETGLITPNKHGLLQNPYRIPPLPIKSFVK---TKKKNLLEENFEHMSPEREM 1014
Db 1349 IHKDETDLFTDQHNIC-----LKLSGQFMKEGNTQIKEDLSDLTFLEVAKAQEACH 1400
QY 1015 GN-ENIPSTVSTISRNIR-----ENVFKEASSNINEVGSTNEVGSSINEIGSSDENI 1068
Db 1401 GNTSNKEDLTATKTEQNIKDFETSDTFFQTASGNISVAKESFNKI---VNFQDQKPEEL 1457
QY 1069 QAEGLNRGRKPLNAMLRLGLVQPEVYKOSLPGSNCKHPEIKKQE-----YEEVQTVNTD 1123
Db 1458 -----HNSLSEL-----HSDIRKNMDILSYEE-----TD 1484
QY 1124 FSPYLISDNLEQPMGSSHASQVCSSETPDDLDDGEIKEDT-----SFAENDIK 1171
Db 1485 IVKHKILKE-SVPVGTGNQLVTFQGPQE---RDEKINEPTLLGFHTASGKKVKIAKESLD 1540
QY 1172 ESSAVFSKSVQKGLSRSPSPFTHLQAGYRCAKAKLESSEENLSSEDEELPCFQHLFF 1231
Db 1541 KVNMLFD--EKEOGTSEITSFSH-----QWAKTLKYRE---ACKDLELAC----- 1580
QY 1232 GKVNNIPQSSTRSHVATECL---SKNTEENLLSLKN-----SLNDC-----SNQ 1273
Db 1581 -----ETIETAPKCKEMQNSLNDKNNLSYIETVVPKLLSONLCQTEENLTKSK 1631
QY 1274 VILAKASQEHLSSET-KCSASLFSQOC--SELED-----LTANTNTOPDFLIGSK 1322
Db 1632 SIFLKVKHVENVEKETAKSPATCYTNQSPYSVIENSALAFYTCGRSKTSVSQTSLSLEAKK 1691

-! CAUTION: REF.6 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO FRAMESHIFTS IN POSITIONS 3782 AND 3811.

-! CAUTION: REF.9 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FOUR FRAMESHIFTS IN POSITIONS 29,1653, 1699 AND 1735.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@sib-sib.ch).

EMBL; AJ131593; CAB40713.1; -
EMBL; AJ019691; BAA78718.1; -
EMBL; AJ010770; CAQ03961.1; -
EMBL; AF026245; AAB83384.1; -
EMBL; AF083037; AD22767.1; -
EMBL; AC004013; ABA96867.1; ALT_FRAME.
EMBL; AF091711; AD39719.1; -
EMBL; AF018346; BAA34523.1; -
EMBL; AC000066; AAC60380.1; ALT_FRAME.
Genew; HGNC:376; AKA39.

DOMAIN	2554	2567	PKA-R11 SUBUNIT BINDING DOMAIN.
Coiled coil; Alternative splicing; Polymorphism.			
MIM: 604001; -			

DOMAIN	164	914	COILED COIL	(POTENTIAL)
DOMAIN	944	1022	COILED COIL	(POTENTIAL)
DOMAIN	1100	1185	COILED COIL	(POTENTIAL)
DOMAIN	1253	1280	COILED COIL	(POTENTIAL)
DOMAIN	1336	1392	COILED COIL	(POTENTIAL)
DOMAIN	1434	1459	COILED COIL	(POTENTIAL)
DOMAIN	1585	1659	COILED COIL	(POTENTIAL)
DOMAIN	1857	2455	COILED COIL	(POTENTIAL)
DOMAIN	2544	2561	COILED COIL	(POTENTIAL)
DOMAIN	2603	2776	COILED COIL	(POTENTIAL)
DOMAIN	3065	3092	COILED COIL	(POTENTIAL)
DOMAIN	3124	3470	COILED COIL	(POTENTIAL)
DOMAIN	3587	3689	COILED COIL	(POTENTIAL)
DOMAIN	3726	3730	POLY-LEU.	

DOMAIN	203	292	GLU-RICH.
DOMAIN	321	1010	GLU-RICH.
DOMAIN	1846	2772	GLU-RICH.
VARSPLIC	17	28	MISSING (IN ISOFORM 2 AND ISOFORM 3).
VARSPLIC	1637	1642	OLQEEI -> LATRD (IN ISOFORM 4).
VARSPLIC	1643	3911	MISSING (IN ISOFORM 4).
VARSPLIC	2175	2182	MISSING (IN ISOFORM 3).
VARSPLIC	2175	2183	STDFKQVE -> Q (IN ISOFORM 6).
VARSPLIC	2895	2907	VGFGEVNCFTSLC -> GSISPELAHSDVOTREICSS (IN ISOFORM 2, ISOFORM 3 AND ISOFORM 6).
VARSPLIC	2895	2948	MISSING (IN ISOFORM 5).
VARSPLIC	3901	3911	STTQFHAGMRR -> ALSLTSTWQHHSARETAPFLFEILSH SLG (IN ISOFORM 6).
VARIANT	1347	1347	K -> KO.

CONFLICT	76	76	E -> Q (IN REF. 3).
CONFLICT	475	475	M -> I (IN REF. 3).
CONFLICT	554	554	E -> G (IN REF. 3).
CONFLICT	638	638	R -> S (IN REF. 3).
CONFLICT	663	663	N -> S (IN REF. 3).
CONFLICT	913	913	H -> N (IN REF. 3).
CONFLICT	956	956	K -> N (IN REF. 3).
CONFLICT	980	982	QXH -> PAP (IN REF. 1 AND 2).
CONFLICT	997	997	Q -> P (IN REF. 1 AND 2).
CONFLICT	1001	1001	Q -> P (IN REF. 1 AND 2).
CONFLICT	1020	1020	N -> D (IN REF. 3).
CONFLICT	1028	1028	V -> E (IN REF. 3).
CONFLICT	1626	1626	R -> P (IN REF. 1 AND 2).
CONFLICT	1703	1703	N -> T (IN REF. 3).
CONFLICT	1707	1707	I -> G (IN REF. 3).
CONFLICT	1802	1802	MISSING (IN REF. 5).
CONFLICT	1843	1843	A -> P (IN REF. 3).
CONFLICT	1956	1956	I -> V (IN REF. 3).

FT	CONFLICT	2027	2027	V -> D (IN REF. 5).
FT	CONFLICT	2157	2158	EI -> HE (IN REF. 7).
FT	CONFLICT	2169	2169	E -> V (IN REF. 3).
FT	CONFLICT	2514	2514	L -> R (IN REF. 3).
FT	CONFLICT	2851	2851	I -> N (IN REF. 8).
FT	CONFLICT	2957	2957	E -> D (IN REF. 3).
FT	CONFLICT	2983	2983	P -> S (IN REF. 3, 7 AND 8).
FT	CONFLICT	3087	3087	Q -> H (IN REF. 3).
FT	CONFLICT	3218	3218	Q -> H (IN REF. 3).
FT	CONFLICT	3307	3309	ESE -> QSQ (IN REF. 3).
FT	CONFLICT	3751	3751	P -> A (IN REF. 3).
FT	CONFLICT	3833	3833	T -> S (IN REF. 3).
SO	SEQUENCE	3911	AA; 453664 MW; 3FB1CB1C1819B47AA CRC64;	

Query Match 3.4%; Score 325; DB 1; Length 3911;
Best Local Similarity 18.0%; Pred. No. 6.3e-06;

Qy	7	R	V	E	V	O	N	V	I	N	A	N	O	K	I	L	E	C	P	I	C	L	E	L	I	K	E	P	V	S	T	K	D	H	I	F	C	K	F	C	M	L	K	L	L	N	O	K	G	P	S	O	C	P	L	C	K	N	66
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:					
Db	1351	Q	O	E	L	S	I	S	S	I	Q	-----	Q	O	L	K	E	T	Q	E	Y	E	A	B	I	H	C	L	Q	R	L	O	A	V	S	T	V	P	P	S	L	P	D	S	V	1401													
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:					
Qy	67	D	I	T	K	S	I	Q	E	S	T	R	F	S	O	L	V	E	L	L	I	I	C	A	F	O	L	D	T	G	L	E	Y	A	N	S	N	F	A	K	E	N	-----	S	P	E	H	L	118										
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:					
Db	1402	V	I	T	E	S	A	Q	T	M	T	P	G	S	V	K	-----	N	I	D	T	I	F	S	G	E	F	G	-----	V	K	E	T	N	I	V	L	L	E	K	Y	O	E	Q	1452														
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:					
Qy	119	K	D	E	V	S	-----	I	I	Q	S	M	-----	G	Y	N	R	A	-----	K	R	L	O	S	E	P	N	S	L	O	E	T	S	L	S	V	O	L	S	N	L	159																	
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:					
Db	1453	E	E	E	V	A	K	V	I	V	S	M	S	T	A	F	A	Q	O	T	E	L	S	R	I	S	G	G	K	E	N	T	A	S	S	K	O	A	H	V	C	O	E	Q	H	Y	N	E	M	L	S	O	D	O	I	G	F	1512	
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:					
Qy	160	G	T	V	R	T	-----	L	R	T	K	O	R	I	O	P	O	K	T	S	V	I	E	L	-----	S	D	S	E	D	T	V	N	K	A	T	Y	C	S	V	G	D	O	E	L	L	Q	-----	206										
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:					
Db	1513	Q	T	P	E	T	V	D	V	A	R	F	K	E	E	F	P	L	S	-----	E	L	G	H	E	K	I	L	L	S	N	S	D	P	H	I	P	E	S	K	O	C	V	L	T	I	S	E	M	F	S	K	D	1568					
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:					
Qy	207	-----	I	T	P	O	G	T	R	D	E	I	-----	S	I	D	S	A	K	A	C	E	S																																				

QY 703 TNAPGSFTKCS--NTSELKEPVNPSLPREEKEKLETVKVSNNADPKDMLSGERVLOT 760
Db 2107 KWVP-RFOPTSEHQTRVEQLAN--HLKEKTKDCSELLSKSKE-QLQRDIOERNEEIEKL 2161
QY 761 ERSVSESSISLVPTDGTQBSISLELVSTLGKAKTBPNCVSOCAAFENPKGLIHGCSK 820
Db 2162 EFRVRELEQALLVSAD--TFQKV--DRKHFGAVEAKEPE--LSLEVOQLAERDAIRKEK 2215
QY 821 DNRNDTEGFKYPLGHEVNHRSRETSIEMSELDQAQYQNTFKVSKROSFALFSPNGAAE 880
Db 2216 EITN-----LEEQLQEPRE-----ELENKNEEVQQLHMQLIOKKESTTRLQ---BLEQ 2261
QY 881 ECATFS-----AHSGLKKQSPKVTPE-----CEQKE-ENQCKNESNIKPQVTYN 924
Db 2262 ENKLFKDDMEKGLAIKESDAMSTQDQVLFGRKFAQIIQKEVEIDQLINEQVTKLQOOLK 2321
QY 925 ITAGFPVVGQKDPVNAKSIKSGSRFCISSQFRGNETGLITPNKKGILLONRIPPLF 984
Db 2322 ITDNKVIEKNELLIRLETOIE-----CLMSD----- 2349
QY 985 PIKSFVTKCKKNLLENFEE-----HSMSPER----- 1012
Db 2350 -----QECVKRNEEIEIQLNEVIEKLOQLANICQKTMNAHSLESEADSLKHOLDV 2402
QY 1013 -----EMGNENIPSTVSTISRNIRENVFK-----EASSSINIEVGS 1049
Db 2403 VIAEKLALEQQVETANEEM-----TFMKNVLKETFNMKNQLTQELFSLKRESEVEKIQS 2457
QY 1050 -STNEVGSSINETGSSDENTQAEGRN-----RQPKNAMLRGLVQ 1090
Db 2458 IPENSVNAIDHLSKDKPELEVLVITDALKSLENQTYFKSFEENGKSGSIINLETRLLQLE 2517
QY 1091 PEVYKQSLPGSNCKHPRIKOEVEEVQTVNTDFPYLISDNLPQPMGSSHAQVCSQETP 1150
Db 2518 STVSAKOLELTOC--YKQKMDQEQGFETEMLQKKIYNLQKIVEKVAALVSOIQLEA- 2575
QY 1151 DDLDDGEIKEDTFAENDIKESSAVFSKSVQ-----KGELSRSPSPFT----- 1194
Db 2576 -----VQYAKFCQDNQTTISPEPTNTQNLNQLREDELGSDISALTLRISELESQV 2627
QY 1195 ---HTHQAQYRRGAKKLESEENLSEDELPFCFHLLFGKVNINFSQSTRHSTVATEC 1251
Db 2628 VEMHTSLI-----LEKEQVTAENKVLKEKRLLEQLQLEKNEKKQREKKRSPQDVEV 2683
QY 1252 LSKNTEENLLSLKNSLNDKSNVITLAKASOBHHLSEETKCSASLFSQCSLELDLTANTN 1311
Db 2684 LKTTTE---LPHSNEESGFFNELEALRAE-----SVATKAELASYKEKAEKLQELLVKE 2735
QY 1312 TQDPFLIGSKQMR-HQSES-QGVGLSDKELVSDDEER----- 1347
Db 2736 TNMTSLQKDLQSVRDHLAAEAKELSLKEKEDTEVQESKKACMFPEPLPKLSKSIAEQTD 2795
QY 1348 GT-GLENNQEBQMSDNLG---EASGCSSETSUSDECSGLSSQ-----SDILTT 1394
Db 2796 GTLKISSNQTPQTLVKNAGIQINLQSECSE-EVTEIISQFTEKIEKMOELHAAEILDM 2854
QY 1395 QORDTMHNLK-----LOQEMAELEAVLE----- 1419
Db 2855 ESRHISETELKREHYAVVOLLKEECOTLKAVIQCLRSKEVFGFYNNCFSTLDSGSDWG 2914
QY 1420 -----QHGSQPSNSYP----- 1430
Db 2915 QGIYLTSHSQGFDIASGRGESESATDFFPKIKGLLRVHNEGQVLSLITESPYSDGED 2974
QY 1431 -----SIIDSSALELURNPEQSTSEKAVITUSQKSE----- 1462
Db 2975 HSIQVSEPWLEERKAVINTISSLKDLITKMLQREAEVYDSSQSHESFSDWRGELLAL 3034
QY 1463 --YPISONPELSADKFEVSADST-----SKNKEPQGVERSSPSKCPSLDDRWYM 1510
Db 3035 QOVFLERSVLLAAFRTELALGTDTDAVGLNLCLEQRIQIOGQVEYQAAMECLQADR--- 3091

RESULT 10

CENP_HUMAN
ID CENP_HUMAN STANDARD; PRT; 3210 AA.
AC P49454; Q13246; Q13171;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE CENP-F kinetochore protein (Centromere protein F) (Mitosis) (AH
DE antigen).
DE CENPF.
GN CENPF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=95348175; PubMed=7542657;
RA Liao H., Winkfein R.J., Mack G., Rattner J.B., Yen T.J.;
RT "CENP-F is a protein of the nuclear matrix that assembles onto
RT kinetochores at late G2 and is rapidly degraded after mitosis.";
RL J. Cell Biol. 130:507-518(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95379848; PubMed=7651420;
RA Zhu X., Mancini M.A., Chang K.-H., Liu C.-Y., Chen C.-F., Shan B.,
RA Jones D., Yang-Feng T.L., Lee W.-H.;
RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein
RT that is specifically involved in mitotic-phase progression.";
RL Mol. Cell. Biol. 15:5017-5029(1995).
RN [3]
RP SEQUENCE OF 2194-3210 FROM N.A.
RX MEDLINE=95336446; PubMed=7612011;
RA Li Q., Ke Y., Kapp J.A., Fertig N., Medsger T.A. Jr., Joshi H.C.;
RT "A novel cell-cycle-dependent 350-kDa nuclear protein: C-terminal
RT domain sufficient for nuclear localization.";
RL Biochem. Biophys. Res. Commun. 212:220-228(1995).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=95370296; PubMed=7642639;
RA Zhu X., Chang K.-H., He D., Mancini M.A., Brinkley W.R., Lee W.-H.;
RT "The C terminus of mitosis is essential for its nuclear localization,
RT centromere/kinetochore targeting, and dimerization.";
RL J. Biol. Chem. 270:19545-19550(1995).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL J. Cell Biol. 143:49-63(1998).
CC -!- FUNCTION: PROBABLY REQUIRED FOR KINETOCHORE FUNCTION, INVOLVED IN
CC CHROMOSOME SEGREGATION DURING MITOSIS. INTERACTS WITH
CC RETINOBLASTOMA PROTEIN (RB), CENP-E AND BUBR1.
CC -!- SUBUNIT: HOMO- OR HETERODIMER.
CC -!- SUBCELLULAR LOCATION: NUCLEAR MATRIX (BUT NOT IN THE NUCLEOLUS),

REORGANIZATION TO THE KINETOCHORE/CENTROMERE (CORONAL SURFACE OF THE OUTER PLATE) AND THE SPINDLE DURING MITOSIS.

-|- DEVELOPMENTAL STAGE: GRADUALLY ACCUMULATES DURING THE CELL CYCLE.

CC -|- PTM: HYPERPHOSPHORYLATED DURING MITOSIS.

CC -----

CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U19769; AAA82889.1; -

DR EMBL; U30872; AAA82935.1; -

DR EMBL; U25725; AAA86889.1; -

DR Genew; HGNC:1857; CENPF.

DR MIM; 600236; -

DR InterPro; IPR001230; Prenyl_site.

KW Chromosomal protein; Nuclear protein; Centromere; Coiled coil;

KW Mitosis; Phosphorylation; Antigen; Cell cycle; Repeat.

FT DOMAIN 14 197

FT DOMAIN 273 769

FT DOMAIN 823 1328

FT DOMAIN 1642 1746

FT DOMAIN 1862 2987

FT DOMAIN 2207 2568

FT REPEAT 2207 2386

FT REPEAT 2389 2568

FT DOMAIN 3015 3032

FT CONFLICT 16 16

FT CONFLICT 250 250

FT CONFLICT 272 272

FT CONFLICT 611 611

FT CONFLICT 1494 1589

FT CONFLICT 1611 1611

FT CONFLICT 1811 1811

FT CONFLICT 2242 2243

FT CONFLICT 2335 2335

FT CONFLICT 2492 2492

FT CONFLICT 2545 2551

FT ELNERNVAALHNDQEAACK -> SSMREWQPCIMTKKPV

FT (IN REF. 3).

SQ SEQUENCE 3210 AA; 367589 MW; 11D83324960E4334 CRC64;

Query Match 3.4%; Score 323.5; DB 1; Length 3210;

Best Local Similarity 19.2%; Pred No. 5.6e-06;

Matches 392; Conservative 290; Mismatches 735; Indels 621; Gaps 93;

QY 9 EEVQNVINAMQ-----KILECPICL-----LIKEPVSTR----- 39

DB 288 QELRNKINLELRLOGHEKEMKGVNKFQLOLEKAKVELIEKVLNKCRLVTRTT 347

QY 40 ---DHIFCFEMLKLLNQKGFSCPLKNDITKRSQESTRFS-----OLVEEL 86

DB 348 AQYDQASTRYTALQELKK-----LTEDLSQRQNAESARCSLEQKIKEKEFEQEL 400

QY 87 LKTIICAFQ-----LDTGLEYA-NSYNFAKKE-----NNSPEHLKD-- 120

DB 401 SROORSFQTLDOECOMKARLTQELQQAKNHNVHVAQELDLTTSVKQOLENLEEFKOKL 460

QY 121 -----EVSIIQSMGYRNRAKRLQSEPEPNSLQETSLSVLSNLGTVRTLRT 167

DB 461 CRAEQAFQASQIKENELRSMEEKNNLLKSHSEQAREVCHLEAKNI----- 512

QY 168 KQRIQPKQTSVIEIGSDSEDTVNKATCVSGDQELQITPQGTDRDISLSAKKAACE 227

DB 513 KOCLNQSQNFABEMAKNFTSQETMLRDQEKINQOE-----NSLTLEKLKLAVAD 562

QY 228 F-----SETDVNTTEHH-OPSNNDINTTEKR-----AERHPKCYOGSSVSNLHVEPC 274

DB 563 LEQRDCSODLLKREHHIEQLNDKLSKTEKESKALLSALELKKYE-----ELK 613

QY 275 GTNTHASSLOHENSLL--LTKDRMNVKE-----AEFCNKSQOPGLARSOHNRWAGSKETC 328

DB 614 BEKTLFCWKSENEKLLTQMESEKENLQSKINHLETCLTKQ--IKSHEYNERVRTLEM- 670

QY 329 NDRRTPTSTB-----KKVDLNAADPLC-----ERKENWKQ-----LPCSENP 364

DB 671 -DRENLSVEIRNLHNVLDKSVSEVETOKLAYMELOQKABFSDQKHQKETEINCLTKTSOLT 729

QY 365 RDTEDVPWITLNSISQKYNWFSDSEDLGLGSDSDHGESENAKVADVLDVLENVDE--Y 422

DB 730 QGVED-----LEHKQLQL-----SNEIMDKRCY-----QDLHAYESLRDLKSKDASLV 775

QY 423 SGSSEKIDLLASDPHEALICKSERVSKSVESNIEDKIFGKTYRKKASLPNLSHVTENLI 482

DB 776 TNEDHORSLLAFDQOPAM-----HHS-----FANIIGECSMPS----- 809

QY 483 IGAFVTEPQIITQERPLTNKLRKRPTSGLHPEDFIKADLAVQKTPMINOGTNOTEQN 542

DB 810 -----ERS-ECRLEADQSPKNSAILQNRVDSLEFSLEQOKMNSDLQKQCE-- 854

QY 543 GOVMNITNSGHENKTKGDSIONE--KNPNPTESLEKE-SAFKTKAEPITSSISNMELEL 598

DB 855 -ELVOIKGEIEENLMAQMHQSFAETSQRIKSLQEDTSAHQNVVAETLSALENKEKEL 913

QY 599 NIHNSKAPKNRLRRKSTRIHALEYVSRNLSPNPCTELQIDSCSSSEEEKKKKYNOM 658

DB 914 QLLNDKV-----ETEQAIEQLKKSNNHLEDSLKLQL--LSETLSLEKKEMSSI 961

QY 659 PYVHRNLO-----LME-----GKBPATGAKSKNPNQETSKRHD-----SOTFPEL 700

DB 962 ISLANKREIEELTQNGTLKEINASLQEKMNLIQKESFANYIDREKSISELSQYKQE 1021

QY 701 KLTNAPGSFTKCSNTSELKEFVNPSL-PREEKEELETV----- 738

DB 1022 KLI-----LLQCEETGNAYEDLSQYKAAQEKNSKLECLLNCTSLCENRKNLEOLKEA 1077

QY 739 -----KVSNNAEADPKDMLSGERVLOQTERSV-----ESSISLVPGTDYGTQESIS 784

DB 1078 FAKHOBELTKLAPAEARNQNLMELETVQOALRSEMTDNQNNKSSEAGGL--KQEIWT 1134

QY 785 LLEVSTLCKAKTEPKNKCVCQAAEFENPKGLIHGC-----SKDNRNDTEGPKYPLGHEVNHIS 840

DB 1135 LKEEO--NKMKEVNDLQENQOLMKVMKTKHECONLESEPIRNSVKE-RESERQCNFK 1191

QY 841 RETSTEMEPESELDQYQLQNTFKVKSRQSFALFSPNPGNAEECATFSAHSGSLKSKSPKVT 900

DB 1192 PQMDLEYKEISLD-----SYNAQLVQLEAMLRNK-----ELKLESEKE 1230

QY 901 FECEQKEENQCKNE--SNIKPVQTVNITAGFPVYVQKDKPV----- 939

DB 1231 KECLQHELOQTIRGDLTSLNLDQMOSQETS-----GLKCEIDAEEKYISGPHELSTSQN 1284

QY 940 DNAK-----CSIKGSRFCLLSSQFRGNETGLITPNK----- 970

DB 1285 DNAHLQCSLQTTMKNKLEIKELQAEKYELVTELNDRSSECTATRKMAEEVQKLLN 1344

QY 971 -----HGLL-----QNPYRIPPLPIKSFVK-TKCKKNL----- 998

DB 1345 EVKILNDDSGLLHGLVELDIPGGEFGEQNPQHVSLAPLDESNSYEHLTLSDKREVQMIF 1404

QY 999 --LEENF-----EEHSMSPER-----EMGNENIPSTVTSIRN----- 1029

DB 1405 AELQEKFLSLQSEHKILHDQHCQSSKMSSELYTVYDSLKAENL--VLSTNLNRFQDGLVK 1462

QY 1030 -----NIRENVEKASSNINEVGSTNEVGSS-----INEIGSSD--ENTQAEIQRNG 1077

DB 1463 EMQGLGLEGLVPSLSSCCVPD-SSLSLSLGDSSFYRALLEQTGDMSSLNLEGAVSANO- 1520

QY 1078 PKLNAMRLGLQPEVYKQSLPGNSKHPETKQBYEEVQTVN-TDFSPYLISD--NL 1133

DB 1521 -----CSVDEVFCSLL-----QTYVDSLKAENLVLSTNLNRFQDGLVKEMQLGL 1564

QY 1134 BQPMGSSSHASOVCSSETPDDLLDGDGEIKEDTSAFENDIKES----- 1173

Db 1165 EGVLPVLSSS-CVPDSSSLSSG-----DSFYRALLEQTGDMSSLNLEGSWSANQCSV 1619
QY 1174 SAVFSKSVQKGLSRSPSPFHT-----HLAQYRRGAKKLESS-----1212
Db 1620 DEVFCSSIQEENLTRKETPSAPKGVLESLCEVYRQSLKLEEKWESOGIMKNKELQE 1679
QY 1213 -EENLSSEDELPFCQHLLFGKVNIPSQST-----RHSTVATECLSKNTEENLLSLKNS 1266
Db 1680 LEQLLSERQELDLRQYLSSENQWQKLTSVTLEMESKLAAB--KKQTEQLSLELEVA 1737
QY 1267 L-----NDCSNQVTLAKSQEHLLSETKCSASFSCQSELEDLTANTQDPFLGSS 1321
Db 1738 RLQQLGLDLSRSULGIDTDAIQGRNESCDS-----KEHTSETTERTP-----1782
QY 1322 QOMRHQSESQVGLSDKELVSD--DEER--GTGLEENNOE---EQSMDSNL---GE---1367
Db 1783 KHDVHQ-----ICDKDAQODLNDIEKITETGAVKPTGECSEQSPDINYPEPPGEDKT 1835
QY 1368 -AASGCSETSVS-----ECSGLSSQSDILTTOOR--DTMQHNL-----IK 1406
Db 1836 QGSSECISELSFGPNALVPMDFLGNQEDIHNLQLRVKTSENLRLLHVEDRDRKVES 1895
QY 1407 LQQEMAELEAVLEQHGQSPNSYPSIISDSSALEDLRNPQSTSEKAVLTSQKSE-YPI 1465
Db 1896 LLENMKLESLKHLQEVQLMTKTEACIELEKIVGELKKNDSUSEKLEYFSCDHQELLQR 1955
QY 1466 SONPEGLSADKFVSADSSTSKKNKEPGVERSSPKCPSLDLRWYMHSCGSLQNRNYPQ 1525
Db 1956 VETSEGLNSD-LEMHADKSSREDIGDNVAK-----VNSW-----KERFLDVE 1997
QY 1526 ELIKV-----VDVEEQLESQGHDLTFETSYLPQDLLEGTPYLESGISLFSDDPESDPE 1581
Db 1998 NELSIRSEKASIEHEALYLEADLEVQVTEKL---CLEKDNENKQKIVIVCLEEELS VVTS 2054
QY 1582 DRAPESARVGNIPSTSAIKVPOLKVAESAQSPAAATTTDTAGNAMEESVSRKEPRL 1639
Db 2055 ERNOLRGELDTMSKTKTALDQLEKMEKTEOE-LESHQSECLHCIOVAAEAQVREKTEL 2111

RESULT 11
YM67_YEAST
ID YM67_YEAST STANDARD; PRT; 1658 AA.
AC Q03661; Q04988;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 187.1 kDa protein in GUA1-ERC8 intergenic region.
GN YMR219W OR YMR261.13 OR YW9959.01.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 1-711 FROM N.A.
RC STRAIN=S288c / AB972;
RA Dedman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
RA Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 608-1648 FROM N.A.
RC STRAIN=S288c / AB972;
RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 249809; CAA89934.1; -.

DR EMBL; 249939; CAA90190.1; -.
DR SGD; S0004832; YMR219W.
KW Hypothetical protein.
SQ SEQUENCE 1658 AA; 187137 MW; 3893F968305A757D CRC64;

Query Match 3.3%; Score 322; DB 1; Length 1658;
Best Local Similarity 19.7%; Pred. No. 2.8e-06;
Matches 344; Conservative 255; Mismatches 704; Indels 444; Gaps 76;

QY 132 NRAKRLLOSPEPNPSLOETSLSVQLSNLGVTRLTKRQKQPOKTSYVIELGSDSSSDTV 191
Db 97 NRKGSILTTLEKEDALFERSLEEE-----RQRFQ-----LHDSLM 131
QY 192 NKATYCSVGQQLLI--TPQGT-----RDEISLDSAKKAACEFSSETDTNTEHIQPSN 243
Db 132 NKYTGNKSQRLLDLRLKSOYQDTSFQNNDEIPLDS-----FISSPLPADESSN 184
QY 244 NDUNTTKRAAERHPEKYQGSSVSNLHVPCGNTTHASSLQHENS-----SLLLTNR- 296
Db 185 IDSQKDEDELEGQ-----SLIKDFLE---NDEVELSEEEKNSGQSSPSIMILSDEE 234
QY 297 -----MNVEKAFCNKSQOPGLARSOHN-----RWAGSKETCNDRRTPSTTEKK 339
Db 235 YABEGALQDVSNDEYAEQGVKRNIGQEQAVNENATQISSDSSSEGQNYSEGVEMELE 294
QY 340 VLDNADPLCERKE-----MNKQRLPCSENPRDTEDPWI-TLNASSIQKNEWF 386
Db 295 DDIDVESDAEKDESOGAEGTEHSVDFSKYMQPRTDNTK----IPVIEKYESDEHKVHQRY 350
QY 387 SR-----SDELLGSDSDHSGESENNAKVDVLVNEVDYSGSSEKI-DLLASDPHE 438
Db 351 SEDGAFDGVSNISVDDESEDEESQAESYSAANAENVVHHNEHEDLDDKELIEDIESD--- 407
QY 439 ALICKSRVHSKSVESNIEDKIFGKYRKKASLPNLSHVTEENLLIGAFVTEPOIIQERPL 498
Db 408 -----SESQAQSEQSGEDDFEYKMKNEKSTSEETNTSES-----RDQGFADKDAYT 455
QY 499 TNKLKRKRRTSGLHPEDFIKKADLAVQKTPMINQCTNOTEQNGOVNMINTNSGHENKT 558
Db 456 KNKVEQGENDE---EPE---KDDIIRSSLDKNFHNKNSKEYSENVL-----ENETD 501
QY 559 GDSIQNEKNPNPIESLEKESAFKTKAETPSSISNMELNIIH-----601
Db 502 PAIVERENQINDVEGYD-----VTGKSVESDLHEHSPDNLDYLAARAMLQ 546
QY 602 -----NSKAPKKNRLRRKSTRIHAIHLVVRNLPPNCTELQIDSCSSSEIKKKYN 656
Db 547 FQOSRNSNCPQKE-----EQVSESYLGHSGNSLNGSLRSLDESE-----585
QY 657 QMPVR-----HSRNLQIMECKEPATGAKKSNKNEQTSKRHSDSTPPELKLTNAPGSFTK 711
Db 586 QIPLKQFTGNNNNLKTDRGLDSSSVEIEVEKYSE---KKLDGSTKEKEL---VPLSTD 638
QY 712 CSNTSEL--KEFVNPSLPREEK--EEKLETVKVYNNNAEDPKDMLMSGERVLOTERSVSS 768
Db 639 TINSSLGNEDSIYSLDDADAISENLTDVPLMEIKTTPKYEIVIESV-YSTSYSEDNT 697
QY 769 ISLVP-----GTDYCTQESISLLEVSTLGLKAKTEPNKCVSQCAAFENPK 812
Db 698 VAMPPQVEYTSPPFMDNPFNSLNDDEYKHKDLKSTLAALAPAFTKKD-----AEEVE-A 750
QY 813 GLTHGGSKNRNDTEGFKYPLGHEVNHRSRETSIEMEESLDAQY-LQNTFKVKSKROS--F 869
Db 751 GVTKSLTSTSGHTNIF-----HTSKETKQVS-DLDESTENVTFENGTGDKNKSQNF 804
QY 870 ALFSNFGNAEEECATPSAHSGSLKKQSPKVTFECEQKEENQGNESNI-----KPVQTVNI 925
Db 805 PGVANSTDKSTEDTDEKYPESAINVTNVTGSSDCSDIETASNVVEENLRYCEKDMNEAEM 864
QY 926 TAGFPVVGQDKPVDNAKSIKGSGRFLCSS--QFR-GNETGLITPNKHGLLONPPIPP 982
Db 865 SSGDECVKQND---DGSKTQIS----FSTDSPDNFQESNDNTEFSSTKYKV-----RNSD 912

Db 4259 PEAKTK 4264

RESULT 13

ATRX_HUMAN
 ID ATRX_HUMAN STANDARD; PRT; 2492 AA.
 AC P46100; P51068; Q15886; Q9NTS3; Q9H0Z1;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked
 nuclear protein) (XNP) (Znf-HX).
 GN ATRX OR RAD54L OR XH2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), VARIANT S-1860, AND
 RP VARIANTS ATR-X.
 RX MEDLINE=97123494; PubMed=8968741;
 RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,
 RA Gibbons R.J.;
 RT "ATRX encodes a novel member of the SNF2 family of proteins: mutations
 RT point to a common mechanism underlying the ATR-X syndrome.";
 RL Hum. Mol. Genet. 5:1899-1907(1996).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
 RX MEDLINE=97386582; PubMed=9244431;
 RA Villard L., Lossi A.-M., Cardoso C., Proud V., Chiaroni P.,
 RA Colleaux L., Schwartz C., Fontes M.;
 RT "Determination of the genomic structure of the XNP/ATRX gene encoding
 RT a potential zinc finger helicase.";
 RL Genomics 43:149-155(1997).
 RN [3]
 RP SEQUENCE OF 860-2492 FROM N.A.
 RX MEDLINE=95179111; PubMed=7874112;
 RA Stayton C.L., Dabovic B., Gulsano M., Gez J., Broccoli V.,
 RA Giovanazzi S., Boscolasco M., Monaco L., Rastan S., Boncinelli E.,
 RA Bianchi M.E., Consalez G.G.;
 RT "Cloning and characterization of a new human Xql3 gene, encoding a
 RT putative helicase.";
 RL Hum. Mol. Genet. 3:1957-1964(1994).
 RN [4]
 RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE=94214473; PubMed=8162050;
 RA Gez J., Pollard H., Consalez G., Villard L., Stayton C.L.,
 RA Millasseau P., Khrestchatsky M., Fontes M.;
 RT "Cloning and expression of the murine homologue of a putative human
 RT X-linked nuclear protein gene closely linked to PGK1 in Xql3.3.";
 RL Hum. Mol. Genet. 3:39-44(1994).
 RN [5]
 RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.
 RX MEDLINE=95211835; PubMed=7697714;
 RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
 RT "Mutations in a putative global transcriptional regulator cause X-
 RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
 RL Cell 80:837-845(1995).
 RN [6]
 RP SEQUENCE OF 1375-2492 FROM N.A.
 RA Pearce A., Chapman J.;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP EZH2 BINDING.
 RX MEDLINE=98167853; PubMed=9499421;
 RA Cardoso C., Timsit S., Villard L., Khrestchatsky M., Fontes M.,
 RA Colleaux L.;
 RT "Specific interaction between the XNP/ATR-X gene product and the SET
 RT domain of the human EZH2 protein.";
 RL Hum. Mol. Genet. 7:679-684(1998).
 RN [8]
 RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
 RP HETEROCHROMATIN.

RX MEDLINE=20040663; PubMed=10570185;
 RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
 RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
 RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
 RT "Localization of a putative transcriptional regulator (ATRX) at
 RT pericentromeric heterochromatin and the short arms of acrocentric
 RT chromosomes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
 RN [9]
 RP DISEASE.
 RX MEDLINE=2013147; PubMed=10751095;
 RA Villard L., Fontes M., Ades L.C., Gez J.;
 RT "Identification of a mutation in the XNP/ATR-X gene in a family
 RT reported as Smith-Fineman-Myers syndrome.";
 RL Am. J. Med. Genet. 91:83-85(2000).
 RN [10]
 RP VARIANT ATR-X SER-1713.
 RX MEDLINE=97196774; PubMed=9043863;
 RA Villard L., Lacombe D., Fontes M.;
 RT "A point mutation in the XNP gene, associated with an ATR-X phenotype
 RT without alpha-thalassemia.";
 RL Eur. J. Hum. Genet. 4:316-320(1996).
 RN [11]
 RP VARIANT JM GLN-2131.
 RX MEDLINE=96224392; PubMed=8630485;
 RA Villard L., Gez J., Mattei J.-F., Fontes M., Saugier-Verber P.,
 RA Munnich A., Lyonnet S.;
 RT "XNP mutation in a large family with Juberg-Marsidi syndrome.";
 RL Nat. Genet. 12:359-360(1996).
 RN [12]
 RP VARIANTS ATR-X.
 RX MEDLINE=97467722; PubMed=9326931;
 RA Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Azenbauer B.,
 RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppier K., Kurosawa K.,
 RA Levin M.D., Masuno M., Neri G., Pierpont M.E., Slaney S.F.,
 RA Higgs D.R.;
 RT "Mutations in transcriptional regulator ATRX establish the functional
 RT significance of a PHD-like domain.";
 RL Nat. Genet. 17:146-148(1997).
 RN [13]
 RP VARIANT ATR-X LEU-246.
 RX MEDLINE=20123082; PubMed=10660327;
 RA Fichera M., Romano C., Castiglia L., Failla P., Ruberto C., Amata S.,
 RA Greco D., Cardoso C., Fontes M., Ragusa A.;
 RT "New mutations in XNP/ATR-X gene: a further contribution to
 RT genotype/phenotype relationship in ATR/X syndrome.";
 RL Hum. Mutat. 12:214-214(1998).
 RN [14]
 RP VARIANT SHS LYS-1742.
 RX MEDLINE=99347960; PubMed=10417298;
 RA Lossi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C.,
 RA Prieto F., Fontes M., Martinez F.;
 RT "Mutation of the XNP/ATR-X gene in a family with severe mental
 RT retardation, spastic paraplegia and skewed pattern of X inactivation:
 RT demonstration that the mutation is involved in the inactivation
 RT bias.";
 RL Am. J. Hum. Genet. 65:558-562(1999).
 RN [15]
 RP VARIANT CWS THR-2050.
 RX MEDLINE=99326061; PubMed=10398237;
 RA Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M.,
 RA Curtis M.;
 RT "Carpenter-Waziri syndrome results from a mutation in XNP.";
 RL Am. J. Med. Genet. 85:249-251(1999).
 RN [16]
 RP VARIANTS ATR-X E-175; 178-V-K-198 DEL; S-190; P-219; L-246 AND C-249.
 RX MEDLINE=99219535; PubMed=10204841;
 RA Villard L., Bonino M.-C., Abidi F., Ragusa A., Beloune J.,
 RA Lossi A.-M., Seaver L., Bonnefont J.-P., Romano C., Fichera M.,
 RA Lacombe D., Hanauer A., Philip N., Schwartz C.E., Fontes M.;
 RT "Evaluation of a mutation screening strategy for sporadic cases of
 RT ATR-X syndrome.";
 RL J. Med. Genet. 36:183-186(1999).

Db 720 TVSIVTSDSPQKLAIILOQLEHKEQBEAWERATSEDSVSVIIVDVKEPSRVDVSCPELEGVE 779
QY 690 KRHSDTFPELKLITNAPGFTKCSNTSELKEFVNPSLPREEKEKLETKVYSNNAEDPK- 748
Db 780 KCSDSQSWEDI---APEITEPCAENRLDTKEEKSVEYEGDLKSGTAETEPVEQSSQPSL 835
QY 749 -----DLMLSGERVLTOTERSVSSSTISIVPGTDYGTQESISL-LEVSTLTKAKTEPNKCV 802
Db 836 PLVRADDPRLRLQLOEQPOTQKTSLSLTDKSMANAKOLSSDAEAQKLGKPSAHASQSF 895
QY 803 SOCAA-----FENPKGLIHGCKDNNDTEGFKYP-----LGH-----EVNHSRETS 844
Db 896 CESSSTPHFTLPK-----EGDIIPLTGATPPLIGHLKLKLEPKRHSTPTIG 941
QY 845 IEMEESELDAQYLQNTFKYKRSQSFALFSN-----PGNAEEECATFSAHS 889
Db 942 I-----SNYPESTIATSDVMSESMVETHDPIILGSGKGDSCAAPDVDDKLCRLMKLV 993
QY 890 GSLUKOSPKVTFECOEKENOQKN-----ESNIKPVQVNTITAGPPVVVGQKD--KPVDN 941
Db 994 PETEASEESLOFNLEKPATGERKNGSTAVAESVSPQKTMVLSLCICEARQENEARSEDP 1053
QY 942 AKCSIKGGSRCFLSSQFRNETGLITPNKHGLLQNYRIPPLFPKSFYKTKCK- ---- 996
Db 1054 PTPPIRGNLLHFPSSQ--GEEKEKLEGDHTIRQSQOPMKPISPVKDPVSPASQKMVIQ 1111
QY 997 -----NLEENFEHNSPSPER-----EMGEN-----IPSTVSTTS 1027
Db 1112 PSSPOGEAMVTDVLEQDEKGRSTNKENPSKALLERSQNNIGTQMECSLRVPETVSAAT 1171
QY 1028 RNNIRENVKEASSNINEVGSTNEVGSSINEIGSSDENIQAELGRNRPKLNAMLRLG 1087
Db 1172 Q-----TIKNVCEQGTSTVD-----QNFGRQDATVQTERGSEKPK----- 1206
QY 1088 VLQPEVYKOSLPGSNCKHPKIKQVEEYVQVNTDFSYLISDNLEQPMGSSHASQVCS 1147
Db 1207 -----VSAQCDDEP-----SLHSQGGEEF-----DMPQPPHG--HVLURHM 1240
QY 1148 ETDPDDLLDGEIKEKEDTSAENDIKESSAVFSKSVQKGELSRSPPTHTHLAQGYRRGAK 1207
Db 1241 RT-----IREVRTLVTRI-----TDVYVYDGTVEVRK 1268
QY 1208 KLESSENLSSDEEULPCQHLLFGKVNIPQSOSTRHSTVATECLSKNTENLLSKNLSL 1267
Db 1269 VTEETEPEPT-----VEQCEQETEVSQOTGGSS 1296
QY 1268 NDCSN-QVILAKASQHHLSEETKCSASLFSQCS-----ELEDLTANTNTQDPFLIGSS 1321
Db 1297 GDLGDISSFSSKASSLHRTSSGTSLSA-MHSSGSGKGAGPLRGKTSGTPEPADFALPSSR 1355
QY 1322 KQMRHQSESQGVGLSDKELVSDDEERGTLGLENNQBEQSMDSNLGEAAGCCESETSVSED 1381
Db 1356 GPGKLSPRKGVSYQTGTPVCEEDGDAGLIRQ-----GKAPVTPR 1396
QY 1382 CSGLSQSQDILTTQQRDT-----MOHNL-----IKLOQEMAELEAVL 1418
Db 1397 GRGRGRPSRTTGTRTETAPVGPLGTIEDISPNLSDPDKSFSRVVPRVPDSTRRTDVGAGA 1456
QY 1419 EOHGSGPSNSYPIISDSSALDLRNPQSTSEKAVLTQSKSEY----- 1463
Db 1457 LRSDSPETPFOAAGPSDGL-DASPGMSFVGLRVKAKWSNGNPFYPSOKITRDVGAGKY 1515
QY 1464 -----PISQNPGLSADKFEYSADSSTSKNKEPG-----VBR 1495
Db 1516 KLLFDGDECDVLGDKDILLCDPIPLDTEVTALSEDEY-FSAGVVKGRKESGELYYSIEK 1574
QY 1496 SPSKCPSLDDRWYMH-SCSGSLQNRN-----YPSQBELIKVVDVEEQLEEE---- 1541
Db 1575 EGQRK-----WYKRMVILSLEQGNRLREQYGLGPEAVTPLTKAADISLDNLVEGKR 1627
QY 1542 -----SGPHDLTETSYLRQDLEGTPLYESGI-----SLFSDDDPSDPSRAPES 1587
Db 1628 KRNSVNSPATTSASSSTTPTTKITITESPRASMGVLGSKRKLITSEERSPAK-RGRKS 1686

QY 1588 ARV-----GNIPSTSTALKVPQLKVAESAQSPAARAAHTTDTAGY----- 1625
Db 1687 ATVKPGAVGAGEFVSPCESGDNTEBPSAL-----EQRGPLPLNKTLFLGYAFLLTM 1738
QY 1626 -NAMESVSREK-PE-LTASTERVNRKMSVMVSGLTPEEFMLVYKARKHHHTLTNLITE 1682
Db 1739 ATTSDKLASRSLPDGPTGSSEE-----EEEFLEIPPENKQY----- 1775
QY 1683 ETHVVMKTDADFVCP-----RTLKYFLGTAGGKVVVSVFWWTQ 1721
Db 1776 --TESQLRAGAGYILEDNFNEAOCNTAYOCLLIADQHCRKTKFLCLASGIPCWSHVWV-- 1831
QY 1722 SIKERKMLNEHDFEVRGDVNVGRNHQGP-----KRARESQRDKIFRGLIEICCCYGPFTN 1774
Db 1832 -----HDSCHANQLQNYRNYLLPAGYSLEEQRIILDWQPRE-----NPFQN 1871
QY 1775 MP-----TDQLEWNVQLC-----GASVYKEIJSFT-----LGTGVHPVIVVOPDAWTE 1819
Db 1872 LKVLVLSVDOQQNFWLELWSEILMTGGAASVKQHHSSAHNKDIALGVDPVVVTDPSC---PA 1928
QY 1820 GFHAIGOMCEAPVVTREWVLDSVALYQC 1847
Db 1929 SVLKCAEALQLPVVSOEWVI-----QC 1950

RESULT 15
ALM1_SCHPO
ID ALM1_SCHPO STANDARD; PRT; 1727 AA.
AC Q9UTK5; O13313; Q9UTU8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Abnormal long morphology protein 1 (Sp8).
GN ALM1 OR SPAC1486.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Guilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Welljons I., Volktaert G., Aert R., Robben J., Grymonprez B.,
RA Welsch C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomycetes pombe";
RL Nature 415:871-880(2002).
RN [2]
RP SEQUENCE OF 495-1727 FROM N.A., AND CHARACTERIZATION.
RC STRAIN=972;

RX MEDLINE=20123449; PubMed=10660053;
 RA Jimenez M., Petit T., Gancedo C., Goday C.;
 RT "The mlt+ gene from Schizosaccharomyces pombe encodes a coiled-coil
 RT protein that associates with the medial region during mitosis.";
 RN Mol. Gen. Genet. 262:921-930(2000).
 [3]
 RP SEQUENCE OF 644-834 FROM N.A.
 RC STRAIN=968 h90;
 RX MEDLINE=2023868; PubMed=10759889;
 RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
 RA Hiraoka Y.;
 RT "Large-scale screening of intracellular protein localization in living
 RT fission yeast cells by the use of a GFP-fusion genomic DNA library.";
 RL Genes Cells 5:169-190(2000).
 CC -!- FUNCTION: AFTER THE ONSET OF MITOSIS, AT MID- TO LATE ANAPHASE,
 CC CO-LOCALIZES WITH THE MEDIAL ACTIN RING. MAY PLAY A ROLE IN
 CC CYTOKINESIS.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AL133357; CAB62414.1;
 DR EMBL: AF010473; AAB65416.1; ALT_INIT.
 DR EMBL: AB028012; BAA87316.1;
 FT Coiled coil.
 KW DOMAIN 57 361 COILED COIL (POTENTIAL).
 FT DOMAIN 443 463 COILED COIL (POTENTIAL).
 FT DOMAIN 542 740 COILED COIL (POTENTIAL).
 FT DOMAIN 804 1106 COILED COIL (POTENTIAL).
 FT DOMAIN 1223 1427 COILED COIL (POTENTIAL).
 FT DOMAIN 1497 1555 COILED COIL (POTENTIAL).
 FT DOMAIN 1601 1664 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1727 AA; 197858 MW; F820BF8D9C132644 CRC64;

 Query Match 3.38; Score 316; DB 1; Length 1727;
 Best Local Similarity 20.18; Pred. No. 5.2e-06;
 Matches 369; Conservative 267; Mismatches 652; Indels 548; Gaps 84;

 QY 7 RVEEYQVNVAMQKILEPICLELIKPEVSTKCDHIFKVCMLKLNQ----- 54
 DB 85 QAEESRLISVLRN-----ENSLKTNLENQNKRFDAITENQSLRRANSELOEQ 134
 QY 55 -KKGSPCLKNIDITKRSIQ-ESTRFSQVVEELLKTIICAFQDITGLEAYANSYNFAKKEN 112
 DB 135 SKIASEQLSIKADQI--EALQNSHLGELQVQSAHQALSIDIE-----ERKKQHMFA---- 183
 QY 113 NSPEHLKDEVSIIQSMGYRNRKARLLQSEP-----ENSLQETSLSVOLSNLGTVRT- 164
 DB 184 SSSSRVKKEEI-LVQEKSAVSALSDIASDHSKVCYKELVSVSRQVDLEKLAGLAQONTE 242
 QY 165 LRTKRIQOPQTSVYIELGSDS-----SEDVTNKATYCSVGDOELLQITPGTDRDEISLDS 220
 DB 243 LNEKIQLEQKRSYSDGNISLKILETDPTSIKELEEVEVETQKRLTALWESKSELOSEV 302
 QY 221 AKKAACEFSETDVTNTEHQPSNN--DLNTEKRAARHPKPKYGGSSVSNLHVPCGTNT 278
 DB 303 AALQKLTSSQSLYNNVTEELNNKQKLLISENSRELQ-EKYD-SVYSELQVYVKNKNT 360
 QY 279 HASS-----LQENSSLLLTLDKRNWVKEAFCNK 307
 DB 361 SVSAGVGLFSLAOKLSAVQNPEFSFTKVYSDNNKLOQKVSLSKLQDLRLTNKFSFCEQ 420
 QY 308 SKQ--PGLARSOH-----NRWAGSKTCNDRRTPTSEKVKVDLNADPLCERKE---- 352
 DB 421 VKQRPVVKQORSEIVRNMYNMFLESLETSNNLT-----KVQAEELSTKMRQECAYL 475
 QY 353 -WNKQKPCSENPRDTEVPWITLNSSIQVNEWFSRS-----DELLGSDDDSHDGE 402

Db 476 QLTASRTQCSDLSEV-----ICLMAELDHLNETKSRNVPATVQVALDE-YAQNPFSTASE 529
 QY 403 SESNAKVADVLVDLNEVDEYSGSEKIDLLASDPHEALICKSERVHKSVSNSIEDKTFG 462
 Db 530 TLNVNELANFSSIKEAVSKTLELREKV-----RALECDVE-IQKQTVQYQISNAVKE 580
 QY 463 KTYRKKASLNLNLSHVTEENLIIGAFVTEPQIIQERPLTNKLRKRRP-----TSLGHPED 516
 Db 581 NSNTLSEQIKNL-----ESELNSSKIKNESLNLNELLKEMLATSRSSILSNSSAGNIDD 636
 QY 517 FIKKAD-----LAVQKTPEMINQNOTEQNGQVMNITNSGHENKTKG 559
 Db 637 KMKSIDESTRELEKNYEVYRNMETAIQESLSKRNQDL-LSEMEALRKELENSKYQOQLST 695
 QY 560 DSQONENKPNPIESLEKE-----SAFKTKABPISSISINMEULE 597
 Db 696 DRLTNAN--NDVEAFKKKEAKELRSINQNLQDIISRODORASKFABELHLVNSLAERLKE 753
 QY 598 LNIHNSKAPKKNRLRRKSSSTRHIALE-LVYSRNLSPNCTELO-IDSCSSSEIKKKK 654
 Db 754 LNA--SKGKD--LRKTRQERLISENCKLLAERELMSLVSQDLQTLNQOQLSDAARKVK 809
 QY 655 YNQMPVHRHNLQMEGKEPATGAKSN-----KPNK 686
 Db 810 FES--EKESLSLSQKLKE--SNEKMSNDLHSLQSKLEKSGIEYSSRIKTMLEKQSLSE 865
 QY 687 QTSKRHDSTFPELKLTNAPG-----SFTKCSNTS----- 716
 Db 866 DNRKLLDNQOMMEIKLQELNGVIELEKQFSTLEAKFTQOKNTSYSEREALLESLSLQ 925
 QY 717 ----ELKEFYVNPSPREEK-----EEKLETYKVSNAEDPKDMLSGERVQLTQERSV 764
 Db 926 SKHTSLESQYNSLRNIEQLOAASKLAEEWVERVKTEYD-----EYRLOTSESL 974
 QY 765 ESSISILVPGTDXGTQBSISLLEVSTLTKAKTEPNKVCQCAAFENPKGLIHGSKDNRN 824
 Db 975 EKNHLKIT-----SLEQRIVILQ-----DEIASSSLRCEN-----ITKDET 1011
 QY 825 DT-----EGFKYPGLGHEVNHVSRETSIEMSESELDQAYLQNTFKVSK-----RQSALFNSP 875
 Db 1012 RVALLLEENKH-LNNELSSH--NAEKQHLKENDYQOQLLLVTDLDRKTRDY----- 1062
 QY 876 GNAEECATFSAHGSLKQSPKVFTECEQKEE-----NOGKNESNIKPVQTVNITAG 928
 Db 1063 ---EKELLRHADARSTLOKLREDTYKALEQVEDLNKELALKAGINESQ-----P 1108
 QY 929 FPVVGQKDPVDNAKCSIKGGSRFCL-----SSQPRGNETGLITPN-----KHGL--LQNP 977
 Db 1109 FP-ISEKEDPLRQEVYVVKQKQAMLLTQLOSSLNLFABEITSPSPDLDSVMKLGSLQNH 1167
 QY 978 Y-RIPPLFPPIKSFVTKCKKNLLEENFEHSMSPEREMGNENIPSTVTSISR--NNIREN 1034
 Db 1168 VKRISKEMEIIIS-----CQQLL---FLE-----NKKLRTVSSNRVADLQRG 1209
 QY 1035 VFKEASSNNINEVGSST-----NEVGSSINEIGSSDENIQAEIGLRNRPKLN---A 1082
 Db 1210 ITEKDVSTSESVOERSNYLNMVALLNESKSLRENLERNEEVITEL-REKIELTKDTLA 1268
 QY 1083 MLRLGLVQPEVYKQS-----LPGSNCKHPETIKKQEEYEVVQTVNTDFSPYLISDNLEQ 1135
 Db 1269 NFRLNKEQLESQLTEKAQVKKLENSN---EYKRNQEIILLSNSSTST----- 1315
 QY 1136 PMGSSSHASQVCE--TPDILLDD-----GEIK---EDTSAENDIKESSAVFSKVQKE 1185
 Db 1316 ---SSDASRLKNELVSKENLIEELNQLGKSELETYKSKSEDLENERAQNSKIEOLE 1372
 QY 1186 LSRSPSPFTTHLAQGYR-----RGAKKLESSENLSEDEEL-----PCFQ-HLLF 1231
 Db 1373 LK-----NTKLAARWTKYEQVNVKSLKHNQIROQLSOKTSELEAKVAECHQLNQL 1425
 QY 1232 GKYNNIPSQSTRH--STVATECLSKNTEENLLSLKNSLNDSCSNQVILAKASOEHLSEET 1289

Db 1426 NKPSATPTATQSEPTSVSLEEFN-STKELSSTQKLSKSEIMD----- 1467
QY 1290 KCSASLFSSQCELEDTANTQDPELIGSSK-----QMRHQSESQGVGLSDKELV 1341
Db 1468 -----ILNTTKKELEKVRQSNKSE-----GTSKDTEIPNEEEMERKKVMQOEVLRLRSRI 1518
QY 1342 SDDEERGTGLEENNQEESQSDNLGEAASGCESETSVS----- 1379
Db 1519 AKELQKNELLRKQNVLOQOVKALQETVVSSEAEASVHADTKDLENLKKTEEMLSVTF 1578
QY 1380 -----EDCSGLSSQSDILTT-----QORDTWOHN----- 1403
Db 1579 QVIFNESISDFSTADFTTFVQKEWEKREIILQKDVEEQVAQSHQKQLDNIRKELEMRN 1638
QY 1404 ---LKLQCEMAELEAVLEQHGQSPNSYPSIISDSSALEDLRNPQESTSEKAVLTQ-- 1458
Db 1639 KLKLSMLEKNLARVRAELEQSKKKDS---PAILS---LEASKNTDSNKSNSVPPAAQVK 1691
QY 1459 -----KSSEYPISONPEGLSADK-FEVSADSSTSK 1487
Db 1692 EKXLIARTHSVDTNSPPKRSSSDAGMDVSNDRKAK 1727

Search completed: January 22, 2003, 16:50:38
Job time : 80.5 secs

GenCore version 5.1.3

Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 16:41:53 ; Search time 68 Seconds

(without alignments)
5645.085 Million cell updates/sec

Title: us-09-734-672-4

Perfect score: 9649

Sequence: 1 MDLSALRVEEVQNVINAMQK.....LYQCQLDTYLIPQIPHSY 1863

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

1: sp-archaea.*

2: sp-bacteria.*

3: sp-fungi.*

4: sp-human.*

5: sp-invertebrate.*

6: sp-mammal.*

7: sp-mhc.*

8: sp-organelle.*

9: sp-phage.*

10: sp-plant.*

11: sp-rodent.*

12: sp-virus.*

13: sp-vertebrate.*

14: sp-unclassified.*

15: sp-rvirus.*

16: sp-bacteriap.*

17: sp-archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9471	98.2	1863	6 Q9GKK8	Q9gkk8 pan troglod
2	9233.5	95.7	1792	4 O15129	O15129 homo sapien
3	5795	60.1	1141	6 O46485	O46485 gorilla gor
4	5776	59.9	1141	6 O46484	O46484 pan troglod
5	5671	58.8	1141	6 O46486	O46486 pongo pygma
6	5443.5	56.4	1140	6 O46487	O46487 macaca mula
7	5014.5	52.0	1146	6 O46488	O46488 alouatta se
8	4895	50.6	1812	11 Q921D2	Q921d2 mus musculus
9	4873.5	50.5	1817	11 Q94952	Q94952 rattus norv
10	4386.5	45.5	1146	6 O46490	O46490 cynocephal
11	4040	41.9	1141	6 O46489	O46489 galago cras
12	3693	38.3	1063	11 Q924E0	Q924e0 pedetes cap
13	3557.5	36.9	947	6 Q8WMT3	Q8wmt3 tapirus pin
14	3555.5	36.8	947	6 Q9BDV3	Q9bdv3 diceros bic
15	3482	36.1	942	6 Q9BDV4	Q9bdv4 equus cabal
16	3479	36.1	973	11 Q924E1	Q924e1 graphiurus

17	3452	35.8	944	6 Q9BDV2	Q9bdv2 lama glama
18	3435	35.6	959	11 Q924E3	Q924e3 sciurus nig
19	3419.5	35.4	963	11 Q99MX4	Q99mx4 glaucomys v
20	3406.5	35.3	931	6 Q9BDW2	Q9bdw2 chaetophrac
21	3382	35.1	948	6 Q9BDU6	Q9bdue felis silve
22	3361	34.8	942	6 Q9BDU8	Q9bdue physeter ca
23	3356	34.8	940	6 Q9BDU7	Q9bdue megaloptera n
24	3345	34.7	759	4 Q92897	Q92897 homo sapien
25	3337.5	34.6	931	6 Q9BDU4	Q9bdue trichechus
26	3336.5	34.6	949	6 Q9TT62	Q9tt62 hipposidero
27	3318	34.4	934	6 Q8SQG2	Q8sqg2 natalus str
28	3298	34.2	946	6 Q9BDV0	Q9bdv0 sus scrofa
29	3294	34.1	938	6 Q9BDW0	Q9bdw0 bradypus tr
30	3293	34.1	924	6 Q9BDU5	Q9bdue dugong dugo
31	3283.5	34.0	945	6 Q9BDW1	Q9bdw1 tamandua te
32	3276	34.0	936	6 Q9TT68	Q9tt68 myotis daub
33	3263	33.8	942	6 Q9BDV5	Q9bdv5 manis sp o
34	3262.5	33.8	931	6 Q8WMT8	Q8wmt8 socialia flu
35	3245	33.6	930	6 Q9BDU3	Q9bdus loxodonta a
36	3236	33.5	906	6 Q8SQF7	Q8sqf7 rhinopoma h
37	3234	33.5	928	6 Q9BDU2	Q9bdue elephas max
38	3226	33.4	928	6 Q9BDU9	Q9bdue hippopotamu
39	3219	33.4	944	6 Q9BDV7	Q9bdv7 scalopus aq
40	3217	33.3	930	6 Q9TT67	Q9tt67 tadarida br
41	3197	33.1	928	6 Q9BDW3	Q9bdw3 dasypus nov
42	3171.5	32.9	917	6 Q8WMT4	Q8wmt4 roussettus a
43	3171	32.9	935	6 Q9TT63	Q9tt63 pteropus ra
44	3167	32.8	930	6 Q8SQG1	Q8sqg1 rhinolophus
45	3157.5	32.7	929	6 Q9BDT4	Q9bdt4 orycteropus

ALIGNMENTS

RESULT 1

Q9GKK8	ID	Q9GKK8	PRELIMINARY;	PRT;	1863 AA.
AC	Q9GKK8;	DT	01-MAR-2001 (Tremblrel. 16, Created)		
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)				
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)				
DE	BRCA1 (Fragment).				
GS	Pan troglodytes (Chimpanzee).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.				
OX	NCBI_TaxID=9598;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=BLOOD;				
RA	Takeda R., Hink R.L., Jogodka C., Walter N.A.R., Messier W.;				
RT	"Positive selection on the human BRCA1 gene may have resulted from				
RT	pressure for prolonged care for infants.";				
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.				
CC	!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.				
DR	EMBL; AF207822; AAG43492.1; -				
DR	InterPro; IPR001357; BRCT.				
DR	InterPro; IPR002378; Brst_cancerI.				
DR	InterPro; IPR001990; Granin.				
DR	InterPro; IPR001841; Znf_ring.				
DR	Pfam; PF00533; BRCT; 2.				
DR	Pfam; PF00097; zf-C3HC4; 1.				
DR	PRINTS; PR00493; BRSTCANCER1.				
DR	SMART; SM00292; BRCT; 2.				
DR	SMART; SM00184; RING; 1.				
DR	PROSITE; PS50172; BRCT; 2.				
DR	PROSITE; PS00422; GRANINS_1; UNKNOWN_1.				
DR	PROSITE; PS00518; ZF_RING_1; 1.				
KW	Zinc-finger.	309			K -> E.
FT	VARIANT	309			S -> G.
FT	VARIANT	590			E -> K.
FT	VARIANT	731			E -> G.
FT	VARIANT	1100			E -> G.

OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=TESTIS;	
RX	MEDLINE=96172829; PubMed=8589721;	
RA	Holt J.T., Thompson M.E., Szabo C., Robinson-Benion C., Arteaga C.L.,	
RA	King M.C., Jensen R.A.;	
RT	"Growth retardation and tumour inhibition by BRCA1.";	
RL	Nat. Genet. 12:298-302(1996).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=TESTIS;	
RA	Holt J.T., Robinson-Benion C.;	
RL	Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF003068; AAB61673.1; -	
DR	InterPro: IPR001357; BRCT.	
DR	InterPro: IPR002378; Brst_cancerI.	
DR	InterPro: IPR001990; Granin.	
DR	pfam; PF00533; BRCT; 2.	
DR	PRINTS; PR00493; BRSTCANCER1.	
DR	SMART; SM00292; BRCT; 2.	
DR	PROSITE; PS01172; BRCT; 2.	
DR	PROSITE; PS00422; GRANINS_1; UNKNOWN_1.	
SQ	SEQUENCE 1792 AA; 199692 MW; OCB9H06C0791A44 CRC64;	
Query Match 95.7%; Score 9233.5; DB 4; Length 1792;		
Best Local Similarity 99.8%; Pred. No. 0;		
Matches 1788; Conservative 0; Mismatches 3; Indels 1; Gaps 1;		
QY	72 SLOESTRFSOLVELLKIICAFOLDTGLEYANSYNAFKAENNSPEHLKDEVSIIQSMGYR 131	
Db		
QY	2 SLOESTRFSOLVELLKIICAFOLDTGLEYANSYNAFKAENNSPEHLKDEVSIIQSMGYR 61	
Db		
QY	132 NRAKRLLOSPEPNSLQETSLVSLNLTGTVRLTKRQIQPKTSYVIELGSDSSEDTV 191	
Db		
QY	62 NRAKRLLOSPEPNSL-ETSLVSLNLTGTVRLTKRQIQPKTSYVIELGSDSSEDTV 120	
Db		
QY	192 NKATYCSVGDOELLQITPQGTREISLDSAKKAACFESETDVTNTEHHQPSNNDLNTEK 251	
Db		
QY	121 NKATYCSVGDOELLQITPQGTREISLDSAKKAACFESETDVTNTEHHQPSNNDLNTEK 180	
QY	252 RAARHPEKYQGSVSNLHVPCGNTNTHASSLQHENSLLLTKDRMNVEKAFCNKSQKP 311	
Db		
QY	181 RAAERHPEKYQGSVSNLHVPCGNTNTHASSLQHENSLLLTKDRMNVEKAFCNKSQKP 240	
QY	312 GLARSOHNRWAGSKETCNDRBTPSTEKKYVDLNADPLCERKEWNKQKLPSENPRDTEVP 371	
Db		
QY	241 GLARSOHNRWAGSKETCNDRBTPSTEKKYVDLNADPLCERKEWNKQKLPSENPRDTEVP 300	
QY	372 WITLNSSIQKVNWFSRSDLLGSDSDHGESESNKAVADVLVDLNEVDYSGSSEKIDL 431	
Db		
QY	301 WITLNSSIQKVNWFSRSDLLGSDSDHGESESNKAVADVLVDLNEVDYSGSSEKIDL 360	
QY	432 LASDPHEALICKSERVHKSVESNIEDKIFGKTYRKASLPNLNLSHVTENLIIGAFVTEPQ 491	
Db		
QY	361 LASDPHEALICKSERVHKSVESNIEDKIFGKTYRKASLPNLNLSHVTENLIIGAFVTEPQ 420	
QY	492 IIOERPLTNKLRKRPRTSLGHPEDIKKADLAVQKTPMINOGTNTQTEONGQVMNITNS 551	
Db		
QY	421 IIOERPLTNKLRKRPRTSLGHPEDIKKADLAVQKTPMINOGTNTQTEONGQVMNITNS 480	
QY	552 GHENKTKGDSIQNEKNPNPTIESLEKESAFKTKAEPITSSSISNMELELNTHNSKAPKNRL 611	
Db		
QY	481 GHENKTKGDSIQNEKNPNPTIESLEKESAFKTKAEPITSSSISNMELELNTHNSKAPKNRL 540	
QY	612 RRKSSRTHIALELVVSRNLSPPNCTELQIDSCSSSEETKKKYQNMPVHRSRNLQIMEG 671	
Db		
QY	541 RRKSSRTHIALELVVSRNLSPPNCTELQIDSCSSSEETKKKYQNMPVHRSRNLQIMEG 600	
QY	672 KEPATGAKSNKNPEQTSKRHSDSTPELKLTNAPGSFTKCSNTSELKEFVNPSLPREK 731	
Db		

Db	601 KEPATGAKSNKNPEQTSKRHSDSTPELKLTNAPGSFTKCSNTSELKEFVNPSLPREK 660	
QY	732 EEKLETVKVSNNAEADPKDMLSGERVLQTERSVESSISILVPCGTGYCTOBSISLLEVSTL 791	
Db		
QY	661 EEKLETVKVSNNAEADPKDMLSGERVLQTERSVESSISILVPCGTGYCTOBSISLLEVSTL 720	
Db		
QY	792 GKAKTEPNKCVSOCAAPENPKGLIHGCSKDNNDTGEFKYPLGHEVNHHSRETSEIEMESE 851	
Db		
QY	721 GKAKTEPNKCVSOCAAPENPKGLIHGCSKDNNDTGEFKYPLGHEVNHHSRETSEIEMESE 780	
Db		
QY	852 LDAQYLQNTFKVSKROSFALEFNPAGNAEPCATFSAHSGSLKKQSPKVFCECQKEENOG 911	
Db		
QY	781 LDAQYLQNTFKVSKROSFALEFNPAGNAEPCATFSAHSGSLKKQSPKVFCECQKEENOG 840	
Db		
QY	912 KNESNIKPVQTVNITAGFPVVGQKDPVDNAKCSIKGGSRFCLSSQFRGNETGLITPNKH 971	
Db		
QY	841 KNESNIKPVQTVNITAGFPVVGQKDPVDNAKCSIKGGSRFCLSSQFRGNETGLITPNKH 900	
Db		
QY	972 GLLQNPYRIPPLPIKSFVKTKCKKNLLEENFEHSMSPEREMGNENIPSTVSTISRNNI 1031	
Db		
QY	901 GLLQNPYRIPPLPIKSFVKTKCKKNLLEENFEHSMSPEREMGNENIPSTVSTISRNNI 960	
Db		
QY	1032 RENNVKEASSNNINEVCSSSTNEVCSSSTNEIGSSDENIQAELGNRGPKNLMLRLGVLP 1091	
Db		
QY	961 RENNVKEASSNNINEVCSSSTNEVCSSSTNEIGSSDENIQAELGNRGPKNLMLRLGVLP 1020	
Db		
QY	1092 EYVKQSLPGSNCKHPEIKKOEYEEVQTVNTDFSPYLISDNLEQPMGSSHASQVSETPD 1151	
Db		
QY	1021 EYVKQSLPGSNCKHPEIKKOEYEEVQTVNTDFSPYLISDNLEQPMGSSHASQVSETPD 1080	
Db		
QY	1152 DLLDDGEIKEDTFAENDIKESSAVFSKVQKGLSRSPFTHTHLAOGYRRGAKKLES 1211	
Db		
QY	1081 DLLDDGEIKEDTFAENDIKESSAVFSKVQKGLSRSPFTHTHLAOGYRRGAKKLES 1140	
Db		
QY	1212 SEENLSSEDEELPCFOHLLPGKVNNIPQSOTRUSTVATECLSKNTENLLSLKNSLDCS 1271	
Db		
QY	1141 SEENLSSEDEELPCFOHLLPGKVNNIPQSOTRUSTVATECLSKNTENLLSLKNSLDCS 1200	
Db		
QY	1272 NQVILAKAQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGSSKQMRHOSQ 1331	
Db		
QY	1201 NQVILAKAQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGSSKQMRHOSQ 1260	
Db		
QY	1332 GVLGSKELVSDDEERGTLGEENNBEOQSMDSNLGAAAGCESETSVSEDCGLSSQSDI 1391	
Db		
QY	1261 GVLGSKELVSDDEERGTLGEENNBEOQSMDSNLGAAAGCESETSVSEDCGLSSQSDI 1320	
Db		
QY	1392 LTTQQRDTWOHNLIKLQOEMAELEAVLEQHSQSPNSYPSIISSDSALEDLRNPQESTSE 1451	
Db		
QY	1321 LTTQQRDTWOHNLIKLQOEMAELEAVLEQHSQSPNSYPSIISSDSALEDLRNPQESTSE 1380	
Db		
QY	1452 KAVLTQSQSEYPISONPEGLSADKFEVSADSTSKNKEPGEVRSRSPSKCPSLDDRWMH 1511	
Db		
QY	1381 KAVLTQSQSEYPISONPEGLSADKFEVSADSTSKNKEPGEVRSRSPSKCPSLDDRWMH 1440	
Db		
QY	1512 SCGSLQNRNYPQOEBELIKVVDVEEQOLESQPHDITETSYLPRQDLEGTPLYESGISLF 1571	
Db		
QY	1441 SCGSLQNRNYPQOEBELIKVVDVEEQOLESQPHDITETSYLPRQDLEGTPLYESGISLF 1500	
Db		
QY	1572 SDPESDPSEDRAPEARSARVGNTPSSTSALKVPOLKVAESNAQSPAAAHTTDTAGYNAMEES 1631	
Db		
QY	1501 SDPESDPSEDRAPEARSARVGNTPSSTSALKVPOLKVAESNAQSPAAAHTTDTAGYNAMEES 1560	
Db		
QY	1632 VSREKPELTASTERVNKRSMVVSGLTPEEFMLVYFARKHHITLNLNTEETHVVMKT 1691	
Db		
QY	1561 VSREKPELTASTERVNKRSMVVSGLTPEEFMLVYFARKHHITLNLNTEETHVVMKT 1620	
Db		
QY	1692 DAEFVVCERTLKYPELGIAGGKVVYSYFWVTQSIKERKMLNEHDFEVRGDDVVVNGRHHQPKR 1751	
Db		
QY	1621 DAEFVVCERTLKYPELGIAGGKVVYSYFWVTQSIKERKMLNEHDFEVRGDDVVVNGRHHQPKR 1680	
Db		
QY	1752 ARESQDRKIFRGLEICCYGFTNMPDQLEWMVQLCGASVVKELSSFTLGTGVHPVVVQ 1811	
Db		
QY	1681 ARESQDRKIFRGLEICCYGFTNMPDQLEWMVQLCGASVVKELSSFTLGTGVHPVVVQ 1740	
Db		

```
QY 1812 PDAWTEGNGFHAIGOMCEAPVVTREWLDSVALYOCQELDTYLIPOIPHSY 1863
DB 1741 PDAWTEGNGFHAIGOMCEAPVVTREWLDSVALYOCQELDTYLIPOIPHSY 1792

RESULT 3
O46485 PRELIMINARY; PRT; 1141 AA.
AC O46485;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Breast cancer type 1 susceptibility protein (Breast and ovarian cancer
DE susceptibility protein) (Fragment).
GN BRCA1.
OS Gorilla gorilla gorilla (Lowland gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
OX NCBI_TaxID=9595;
RN [1]
RP SEQUENCE FROM N.A. PubMed=9462745;
RX MEDLINE=98122577;
RA Hacia J.G., Makalowski W., Edgemon K., Erdos M.R., Robbins C.M.,
RA Fodor S.P.A., Brody L.C., Collins F.S.;
RT "Evolutionary sequence comparisons using high-density oligonucleotide
RT arrays."
RL Nat. Genet. 18:155-158(1998).
CC -!- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. MAY PLAY A ROLE
CC IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC EMBL; AF019076; AAC39584.1; -.
DR InterPro: IPR002378; Brst.CancerI.
DR PRINTS: PR00493; BRSTCANCER1.
KW DNA-binding; Nuclear protein; Anti-oncogene.
FT NON_TER 1
FT DOMAIN 277 284 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 427 430 POLY-LYS.
FT DOMAIN 985 1006 LEUCINE-ZIPPER (POTENTIAL).
FT NON_TER 1141 1141
SQ SEQUENCE 1141 AA; 127339 MW; A09531803BB99C3D CRC64;

Query Match 60.1%; Score 5795; DB 6; Length 1141;
Best Local Similarity 98.1%; Pred. No. 3.8e-291;
Matches 1119; Conservative 6; Mismatches 16; Indels 0; Gaps 0;

QY 225 ACEFSETDVTNTEHQPSNNDLNTTEKRAAERHPEKYGQSSVSNLHVPCGTNTNTHASSLQ 284
DB 1 ACEFSERDVTNTEHQPSNNDLNTTEKRATERHPEKYGQSSVSNLHVPCGTNTNTHASSLQ 60

QY 285 HENSLLLTKDRMNYEAFECNKKOPGLARSOHNRWAGSKETCNDRTPTTEKKVDLNA 344
DB 61 HENSLLLTKDRMNYEAFECNKKOPGLARSOHNRWAGSKETCNDRTPTTEKKVDLNA 120

QY 345 DPLCERKENNKOKLPCSENPRDTEDPVITLNSSIQKYNFWSRDELGGSDSDHGESE 404
DB 121 DPLCERNENWKOKLPCSENPRDTEDPVITLNSSTQKYNFWSRDELGGSDSDHGESE 180

QY 405 SNAKYADVLDVNLNVEDEYSGSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKT 464
DB 181 SNAKYADVLDVNLNVEDEYSGSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKT 240

QY 465 YRKASLPNLSHVNTENLIIGAFVTEPQIIQERPLTNLKKRRRPTSLGLHPEDFIKKADLA 524
DB 241 YRKASLPNLSHVNTENLIIGAFVTEPQIIQERPLTNLKKRRRATSLGLHPEDFIKKADLA 300

QY 525 VOKTPEMINQGTNQEQGVNMTNSNGHENKTGDSITONEKPNPTIESLEKESAFKTKA 584
DB 301 VOKTPEMINQGTNQEQGVNMTNSNGHENKTGDSITONEKPNPTIESLEKESAFKTKA 360

QY 585 EPISSISNMELELNHNKAPKKNRLRRKSTRIHHALELVVSRNLSPPNCTELQIDSC 644
DB 361 EPISSISNMELELNHNKAPKKNRLRRKSTRIHHALELVVSRNLSPPNCTELQIDSC 420
```

```
QY 645 SSSEIKKKYNQMPVHRHNRNLQLMEGKEPATGAKKSNKPNEOTSKRHDSDFPELKLTN 704
DB 421 SSSEIKKKYNQMPVHRHNRNLQLMEDKEPATGAKKSNKPNEOTSKRHDSDFPELKLTN 480

QY 705 APGSFTKCSNTSELKEFVNPSLPREEKEEKLTVKVSNNAEADPKDMLSGERVLQTSRV 764
DB 481 APGSFTNCSNTSELKEFVNPSLPREEKEEKLTVKVSNNAEADPKDMLSGERVLQTSRV 540

QY 765 ESSISLVPGTDYGTQBSISLLEVTSLGAKTEPNKCVSQCAAFENPKGLIHGCSKDNRN 824
DB 541 ESSISLVPGTDYGTQBSISLLEVTSLGAKTEPNKCVSQCAAFENPKGLIHGCSKDNRN 600

QY 825 DTEGFKYPLGHEVHNHRSRTSEMESELDADYQNTFKVSKRSFALSPNCAEEECAT 884
DB 601 DTEGFKYPLGHEVHNHRSRTSEMESELDADYQNTFKVSKRSFALSPNCAEEECAT 560

QY 885 FSAHSGSLKKOSPKVTFECEQKEENQGNKSNIRKPVQTVNITAGFPVYQKQDKPVDNAKC 944
DB 661 FSAHRSRLKKOSPKVTFECEQKEENQGNKSNIRKPVQTVNITAGFPVYQKQDKPVDYAKC 720

QY 945 SIKGSRFCLSSQFRGNETGLITPNKHGGLQNPYRIPPLFPIKSFVKTKCKKNLLEENFE 1004
DB 721 SIKGSRFCLSSQFRGNETGLITPNKHGGLQNPYRIPPLFPIKSFVKTKCKKNLLEENFE 780

QY 1005 EHSMSPEREMGNENIPSTVSTISRNNIRENVFKEAASSNINEVSGSSTNEVSGSSINEIGSS 1064
DB 781 EHSMSPEREMGNENIPSTVSTISRNNIRENVFKEAASSNINEVSGSSTNEVSGSSINEVGS 840

QY 1065 DENTQAEELGRNRGPKLNAMLRGVLQPEVYKQSLPGSNCKHPEIKKQYEEVYQTVNTDF 1124
DB 841 DENTQAEELGRNRGPKLNAMLRGVLQPEVYKQSLPGSNCKHPEIKKQYEEVYQTVNTDF 900

QY 1125 SPYLISDNLEQPMGSSHASQVCSPTDLDLDDGEIKEDTSAENDIKESSAVFSKVQKG 1184
DB 901 SPYLISDNLEQPMGSSHASQVCSPTDLDLDDGEIKEDTSAENDIKESSAVFSKVQKG 960

QY 1185 ELSRSPSPFTHTLAQGYRRGAKKLESSEENLSSEDELPCHLFGKVNINPSQSTRH 1244
DB 961 ELSRSPSPFTHTLAQGYRRGAKKLESSEENLSSEDELPCHLFGKVNINPSQSTRH 1020

QY 1245 STVATECLSKNTEENLLSLKNSLNDCSNOVLAKASQEHHLSEETKCSASFSSQCCSELE 1304
DB 1021 STVATECLSKNTEENLLSLKNSLNDCSNOVLAKTQSEHHLSEETKCSASFSSQCCSELE 1080

QY 1305 DLTANTNTQDFPLIGSSQKMRHQSQCQVGLSDKELVSDDEERGCTGLENNQEEQSDMSN 1364
DB 1081 DLTANTNTQDFPLIGSSQKMRHQSQCQVGLSDKELVSDDEERGCTGLENNQEEQSDMSN 1140

QY 1365 L 1365
DB 1141 L 1141

RESULT 4
O46484 PRELIMINARY; PRT; 1141 AA.
AC O46484;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Breast cancer type 1 susceptibility protein (Breast and ovarian cancer
DE susceptibility protein) (Fragment).
GN BRCA1.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98122577; PubMed=9462745;
RA Hacia J.G., Makalowski W., Edgemon K., Erdos M.R., Robbins C.M.,
RA Fodor S.P.A., Brody L.C., Collins F.S.;
```


RT "Evolutionary sequence comparisons using high-density oligonucleotide
RL arrays.";
RL Nat. Genet. 18:155-158(1998).
CC -!- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. MAY PLAY A ROLE
CC IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
DR EMBL: AF019075; AAC39583.1; -.
DR InterPro: IPR002378; Brst_cancer1.
DR PRINTS: PR00493; BRSTCANCER1.
KW DNA-binding; Nuclear protein; Anti-oncogene.
FT NON_TER 1 1
FT DOMAIN 277 284 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 427 430 POLY-LYS.
FT DOMAIN 505 509 POLY-GLU.
FT DOMAIN 985 1006 LEUCINE-ZIPPER (POTENTIAL).
FT NON_TER 1141 1141
SQ SEQUENCE 1141 AA; 127399 MW; D72660B3CA6632D5 CRC64;

Query Match 59.9%; Score 5776; DB 6; Length 1141;
Best Local Similarity 97.9%; Pred. No. 3.6e-290;
Matches 1117; Conservative 6; Mismatches 18; Indels 0; Gaps 0;

QY 225 ACFFSETDVTNTEHHOPSNNDLNTTEKRAAERHPEKYQGVSSVNLHVEPCGTNTHASSLQ 284
Db 1 ACFFSETDVTNTEHHOPSNNDLNTTEKRAERHPEKYQGVSSVNLHVEPCGTNTHASSLQ 60

QY 285 HENSLLLTKDRMNVKAEFCNKSQPLARSQHNWAGSKETCNDRTPTSTKTKVDLNA 344
Db 61 HENSLLLTKDRMNVKAEFCNKSQPLARSQHNWAGSKETCNDRTPTSTKTKVDLNA 120

QY 345 DPLCKERKNKQKLPCCSENPRDTEVPWITLNSSIQKVNEFWSRSDLLGSDSDSHDGESE 404
Db 121 DPLCKERKNKQKLPCCSENPRDTEVPWITLNSSIQKVNEFWSRSDLLGSDSDSHDGESE 180

QY 405 SNAKVADVLVDLVEVDEYSSGSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKT 464
Db 181 SNAKVADVLVDLVEVDEYSSGSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKT 240

QY 465 YRKASLPNLSHVTENLIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPDEFIKRADLA 524
Db 241 YRKASLPNLSHVTENLIIGAFVTEPQIIQERPLTNKLRKRATSLGHPDEFIKRADLA 300

QY 525 VQKTPWINGOTNQTONGOVNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKA 584
Db 301 VQKTPWINGOTNMQEQGVNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKA 360

QY 585 EPISSSISSNLELNTNHSKAPKKNLRKRKSTRHIALELVVSRNLSPPNCTELQIDSC 644
Db 361 EPISSSISSNLELNTNHSKAPKKNLRKRKSTRHIALELVVSRNLSPPNCTELQIDSC 420

QY 645 SSSEIEIKKKYNQMPVHRHSNQLMEGKEPATGAKKSNKPNQTSKRHSDTPELKLTN 704
Db 421 SSSEIEIKKKYNQMPVHRHSNQLMEGKEPATGVKKSNEQTSKRHSDTPELKLTN 480

QY 705 APGSFTKCSNTSELKFVNPDLPREKEEKLFTVKVYVSNNAEDPKDMLSGSERVLQTSRV 764
Db 481 APGSFTKCSNTSELKFVNPDLPREKEEKLFTVKVYVSNNAEDPKDMLSGSERVLQTSRV 540

QY 765 ESSISIVPGTDTGTOESISLLEVTSLGKAKTEPNKCVSOCAAFENPKGLIHGCSKDRN 824
Db 541 ESSISIVPGTDTGTOESISLLEVTSLGKAKTEPNKCVSOCAAFENPKGLIHGCSKDRN 600

QY 825 DTGFKYPLGHEVNHRSRETSIEMEESLDAQYLQNTFKVSKROSFALFSPNGNAEECAT 884
Db 601 DTGFKYPLGHEVNHRSRETSIEMEESLDAQYLQNTFKVSKROSFALFSPNGNPRECAT 660

QY 885 FSAHSGSLKQSPKVFTECEQKEENQKNEBSNTPKPVQVGNKPVQKOPVDNAKC 944
Db 661 FSAHSGSLKQSPKVFTECEQKEENQKNEBSNTPKPVQVGNKPVQKOPVDNAKC 720

QY 945 SIKGSRFCLSSQFRNETGLITPNKHGLLQNPYRIPPLPIKSFVTKCKKNLLENPE 1004
Db 721 SIKGSRFCLSSQFRNETGLITPNKHGLLQNPYRIPPLPIKSFVTKCKKNLLENPE 780

QY 1005 EHSMSPEREMGNENIPSTVTSIRNNIRENVFKFEASSSNINEVGSSTNEVGSSINEIGSS 1064
Db 781 EHSMSPEREMGNENIPSTVTSIRNNIRENVFKFEASSSNINEVGSSTNEVGSSINEVGSS 840

QY 1065 DENIQAEELGRNRPKLNAMRLGLVQPEVYKQSLPDSNCKHPKIKQIEYEVVQTVNTDF 1124
Db 841 DENIQAEELGRNRPKLNAMRLGLVQPEVYKQSLPDSNCKHPKIKQIEYEVVQTVNTDF 900

QY 1125 SPYLISDNLQPMGSSHASOVCSPTDDLLDDGEIKEDTSFAENDIKESSAVFSKSVQKG 1184
Db 901 SPCLISDNLQPMGSSHASOVCSPTDDLLDDGEIKEDTSFAENDIKESSAVFSKSVQKG 960

QY 1185 ELSRSPSPFTHTHLAQGYRGAKKLESSEENLSSEDEELPCFQHLFLGKVNIPSQSTRH 1244
Db 961 ELSRSPSPFTHTHLAQGYRGAKKLESSEENLSSEDEELPCFQHLFLGKVNIPSQSTRH 1020

QY 1245 STVATECLSKNTEENLLSKNSLNDSCNQVILAKASQEHHLSEETKCSASLFSQSCSELE 1304
Db 1021 STVATECLSKNTEENLLSKNSLNDSCNQVILAKASQEHHLSEETKCSASLFSQSCSELE 1080

QY 1305 DLTANTNTODPFLIGSSKOMRHOSESQGVGLSDKELVSDDEERGTLGLEENNOEQSDSN 1364
Db 1081 DLTANTNTODPFLIGSSKOMRHOSESQGVGLSDKELVSDDEERGTLGLEENNOEQSDSN 1140

QY 1365 L 1365
Db 1141 L 1141

RESULT 5
O46486 PRELIMINARY; PRT; 1141 AA.
AC O46486;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Breast cancer type 1 susceptibility protein (Breast and ovarian cancer
DE susceptibility protein) (Fragment).
GN BRCA1.
OS Pongo. pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX NCBI_Taxid=9600;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98122577; PubMed=9462745;
RA Hacia J.G., Makalowski W., Edgemon K., Erdos M.R., Robbins C.M.,
RA Fodor S.P.A., Brody L.C., Collins F.S.;
RT "Evolutionary sequence comparisons using high-density oligonucleotide
RT arrays.";
RL Nat. Genet. 18:155-158(1998).
CC -!- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. MAY PLAY A ROLE
CC IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
DR EMBL: AF019077; AAC39585.1; -.
DR InterPro: IPR002378; Brst_cancer1.
DR PRINTS: PR00493; BRSTCANCER1.
KW DNA-binding; Nuclear protein; Anti-oncogene.
FT NON_TER 1 1
FT DOMAIN 277 284 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 427 430 POLY-LYS.
FT DOMAIN 985 1006 LEUCINE-ZIPPER (POTENTIAL).
FT NON_TER 1141 1141
SQ SEQUENCE 1141 AA; 127450 MW; A4C692E5779091F1 CRC64;

Query Match 58.8%; Score 5671; DB 6; Length 1141;
Best Local Similarity 96.3%; Pred. No. 9.6e-285;
Matches 1099; Conservative 11; Mismatches 31; Indels 0; Gaps 0;

QY 225 ACFFSETDVTNTEHHOPSNNDLNTTEKRAAERHPEKYQGVSSVNLHVEPCGTNTHASSLQ 284
Db 1 ACFFSETDVTNTEHHOPSNNDLNTTEKRAERHPEKYQGVSSVNLHVEPCGTNTHASSLQ 60

QY 285 HENSLLLTKDRMVEKAFCKNSKOPGLARSOHNRWAGSKETCNDRTPTSTTEKKVDLNA 344
DB 61 HENSLLLTKDRMVEKAFCKNSKOPGLARSOHNRWAGSKETCNDRTPTSTTEKKVDLNA 120
QY 345 DPLCKERKNKQKLPCCSENPRDTEVPWITLNSSTQKVNEFWSRDELLGSDSDHGGSE 404
DB 121 DPLCKERKNKQKLPCCSENPRDTEVPWITLNSSTQKVNEFWSRDELLGSDSDHGRSE 180
QY 405 SNAKVADVLVDVNEVDEYSGSEKIDLLASDPHEALICKSERVHKSVESNEDKIFGKT 464
DB 181 SNAKVADVLVDVNEVDEYSGSEKIDLLASDPHEALICKSERVHKSVESNEDKIFGKT 240
QY 465 YRKASLPNLSHVTEENLIIGAFVTEPOIIQERPLTNKLRKRRTTSGLHPDEFIKKADLA 524
DB 241 YRKASLPNLSHVTEENLIIGAFVTEPOIIQERPLTNKLRKRRTTSGLHPDEFIKKADLA 300
QY 525 VQKTPMINOQTNGTONGOVNINSGHENKTGKDSIQNEKNPNPIESLEKESAFKTKA 584
DB 301 VQKTPMINOQTNGTONGOVNINSGHENKTGKDSIQNEKNPNPIESLEKESAFKTKA 360
QY 585 EPISSSISNMELELNHNSKAPKNLRLKRSSTRHIALELVVSRNLPNPPNCTELQIDSC 644
DB 361 EPISSSISNMELELNHNSKAPKNLRLKRSSTRHIALELVVSRNLPNPPNCTELQIDSC 420
QY 645 SSSEEEKKKYKNQMPVHRNRLQMEGEPATGAKKSNKPNEOTSKRHSDFPFELKLTN 704
DB 421 SSSEEEKKKYKNQMPVHRNRLQMEGEPATGAKKSNKPNEOTSKRHSDFPFELKLTN 480
QY 705 APGSFTKCSNTSELKEFVNPSLPREEKEKLETVKVSNNAPDKDMLSGERVLOQTSRV 764
DB 481 APGSFTKCSNTSELKEFVNPSLPREEKEKLETVKVSNNAPDKDMLSGERVLOQTSRV 540
QY 765 ESSSISLVPGTDYGTQESTLSLEVTGLKAKTEPNKCVSOCAAFENPKGLHGCCKNRRN 824
DB 541 ESSSISLVPGTDYGTQESTLSLEVTGLKAKTEPNKCVSOCAAFENPKGLHGCCKNRRN 600
QY 825 DTGFGFYPLGHEVNSRSTSIEMEESELDAQYQNTFVKVRSQSFALPSNPGNAEEECAT 884
DB 601 DTGFGFYPLGHEVNSRSTSIEMEESELDAQYQNTFVKVRSQSFALPSNPGNAEEECAT 660
QY 885 FSAHSGSLKKQSPKVTFECEQKEENOGKNESNIKPVQTNITAGFPVVGQKDKPDVNAKC 944
DB 661 FSAHSGSLKKQSPKVTFECEQKEENOGKNESNIKPVQTNITAGFPVVGQKDKPDVNAKC 720
QY 945 SIKGSRFCLSSQFRGNETGLITPNKHLQNPYRIPPLFPKISPVKTKCKKNLLEENFE 1004
DB 721 SIKGSRFCLSSQFRGNETGLITPNKHLQNPYRIPPLFPKISPVKTKCKKNLLEENFE 780
QY 1005 EHSMSPEREMGNENIPSTVSTISRNIRENVFKEASSNINEVGSSTNEVGSSINEIGSS 1064
DB 781 EHSMSPEREMGNENIPSTVSTISRNIRENVFKEASSNINEVGSSTNEVGSSINEIGSS 840
QY 1065 DENIAELGRNRPKLNAMLRGLVLOPEYVYKQSLPGSNCKHPEIKKQBYEVLQVNTDF 1124
DB 841 DENIAELGRNRPKLNAMLRGLVLOPEYVYKQSLPGSNCKHPEIKKQBYEVLQVNTDF 900
QY 1125 SPYLISDNLEQPMGSSHASQVCEPDDLLDGEIKEDTSAENDIKESSAVFSKSVQK 1184
DB 901 SPYLISDNLEQPMGSSHASQVCEPDDLLDGEIKEDTSAENDIKESSAVFSKSVQK 960
QY 1185 ELSRSPSPPTHPLAQGYRRGAKKLESSEENLSSDEDELPCFQHLGLFGKVNIPQSSTRH 1244
DB 961 ELSRSPSPPTHPLAQGYRRGAKKLESSEENLSSDEDELPCFQHLGLFGKVNIPQSSTRH 1020
QY 1245 STVATECLSKNTEENLLSKNSLNDQSNQVILAKASQEBHILSEETKCSASLFSQCSSELE 1304
DB 1021 STVATECLSKNTEENLLSKNSLNDQSNQVILAKASQEBHILSEETKCSASLFSQCSSELE 1080
QY 1305 DLTANTNTODPPLIGSSKOMRHQSOQVGLSDKELVSDDEERGTLGLENNOEQSGSDSN 1364
DB 1081 DLTANTNTODPPLIGSSKOMRHQSOQVGLSDKELVSDDEERGTLGLENNOEQSGSDSN 1140

QY 1365 L 1365
DB 1141 L 1141
RESULT 6
O46487
ID O46487 PRELIMINARY; PRT: 1140 AA.
AC O46487; O28525;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Breast cancer type 1 susceptibility protein (Breast and ovarian cancer
DE susceptibility protein) (Fragment).
GN BRCA1.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9812577; PubMed=9462745;
RA Hacia J.G., Makalowski W., Edgemon K., Erdos M.R., Robbins C.M.,
RA Fodor S.P.A., Brody L.C., Collins F.S.;
RT "Evolutionary sequence comparisons using high-density oligonucleotide
RT arrays.";
RL Nat. Genet. 18:155-158(1998).
RN [2]
RP SEQUENCE OF 964-1028 FROM N.A.
RA Thompson M.E., Holt J.T.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. MAY PLAY A ROLE
CC -!- IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
DR EMBL; AF019078; AAC39586.1; -;
DR EMBL; U44730; AAB03212.1; -;
DR InterPro: IPR002378; Brst_cancer1.
DR PRINTS: PR00493; BRSTCANCER1.
KW DNA-binding; Nuclear protein; Anti-oncogene.
FT NON_TER 1
FT DOMAIN 276 283 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 984 1005 LEUCINE-ZIPPER (POTENTIAL).
FT CONFLICT 978 978 R -> Q (IN REF. 2).
FT NON_TER 1140 1140
SQ SEQUENCE 1140 AA; 126958 MW; EC1873CAFA5E32F7 CRC64;
Query Match 56.4%; Score 5443.5; DB 6; Length 1140;
Best Local Similarity 92.6%; Pred. No. 5.4e-273;
Matches 1057; Conservative 35; Mismatches 47; Indels 3; Gaps 3;
QY 225 ACEFSETDVTNTEHHQPSNNDLNTTEKRAERHPKPYOGSSVSNLHVPCGTNTTHASSIQ 284
DB 1 ACEFSEKIDITNTEHHQSSNNDLNTTEKHATERHPKPYOGSSVSNLHVPCGTNTTHASSIQ 60
QY 285 HENSLLLTKDRMVEKAFCKNSKOPGLARSOHNRWAGSKETCNDRTPTSTTEKKVDLNA 344
DB 61 HEN-SLLTKDRMVEKAFCKNSKOPGLARSOHNRWAGSKETCNDRTPTSTTEKKVDLNA 119
QY 345 DPLCKERKNKQKLPCCSENPRDTEVPWITLNSSTQKVNEFWSRDELLGSDSDHGGSE 404
DB 120 NALYERKNWQKLPCCSENPRDTEVPWITLNSSTQKVNEFWSRDELLGSDSDHGGSE 179
QY 405 SNAKVADVLVDVNEVDEYSGSEKIDLLASDPHEALICKSERVHKSVESNEDKIFGKT 464
DB 180 SNAKVADVLVDVNEVDEYSGSEKIDLLASDPHEALICKSERVHKSVESNEDKIFGKT 239
QY 465 YRKASLPNLSHVTEENLIIGAFVTEPOIIQERPLTNKLRKRRTTSGLHPDEFIKKADLA 524
DB 240 YRKASLPNLSHVTEENLIIGAFVTEPOIIQERPLTNKLRKRRTTSGLHPDEFIKKADLA 299
QY 525 VQKTPMINOQTNGTONGOVNINSGHENKTGKDSIQNEKNPNPIESLEKESAFKTKA 584
DB 301 VQKTPMINOQTNGTONGOVNINSGHENKTGKDSIQNEKNPNPIESLEKESAFKTKA 360

Db	300	VQKTEIINOQTNOEQNGOVMNITNSAHENKTKGDSIQNEKNPNPIESLEESAFKTKA	359
QY	585	EPISSSISNMELELNTHNSKAPKKNLRKRKSSSTRHHALELVVSRNLSPPNCTELQIDSC	644
Db	360	EPISSSINNMELELNTHNSKAPKKNLRKRKSSSTRHHALELVVSRNLSPPNCTELQIDSC	419
QY	645	SSSEETKKKKYNOMPVRHRSRNLMQMEGKEPATGAKSKNKPNEQTSKRHSDTTPPELKLTN	704
Db	420	SSSEETKKKKYNOMPVRHRSRNLMQMEGKESATGAKSKNKPNEQTSKRHASDTTPPELKLTK	479
QY	705	APGSFTKCSNTSELKEFVNPSLPREEKEKLETVKVSNNNAEDPKDMLSGERVLOQERSV	764
Db	480	VPGSFTNCSTSE-KGFVNPSLPREEKEKLETVKVSNNNAEDPKDMLSGERVLOQERSV	538
QY	765	ESSISILVPGTDYGTQESISLLEVSSTLGAKTEPNKVCVSOCAAFENPKGLIHGCSKDNRN	824
Db	539	ESSISILVPGTDYGTQESISLLEVSSTLGAKTERNKMCSOCAFENPKELIHGCSSETRN	598
QY	825	DTEGFKYPLGHEVNHHSRETSIEMEESELDQYLONTFKYVKRQSFALFSPNPGNAEECAT	884
Db	599	DTEGFKYPLGSEVNHHSQETSIEIESESELDTQYLONTFKYVKRQSFALFSPNPGNPEBCAT	658
QY	885	FSAHSGSLKKQSPKVFECQKKEENCKNESNIKPVQTVNITAGFPVVGOKDPVDNAKC	944
Db	659	FSAHSGSLKKQSPKVFSECEQKEENCKQKOSNIKPVQTVNITAGFVVGOKDPVDNAKC	718
QY	945	SIKGSRFCLSSQFRNETGLITPNKHGILLQNPYRIPPLFPPIKSFYKTKCKNLLLENFE	1004
Db	719	SIKGSRFCLSSQFRNETGLITPNKHGILLQNPYHIPPLFPVKSFYKTKCKNLLLENSE	778
QY	1005	EHSMSPEREMGNEN-IPSTVSTISRNNIRENVPEKASSSINIEVGSSTNEVGSSINEIGS	1063
Db	779	EHSVSPERAVGNENIIPSTVSTISHNNIRENAFKEASSSINIEVGSSTNEVGSSINEVGS	838
QY	1064	SDENIOAELGRNRPKLNAMRLGLVQPEVYKOSLPCNSCKHPKIKKOEYEEVQTVNTD	1123
Db	839	SDENIOAELGRNRPKLNAMRLGLVQPEVCOSLPCISNCKHPKIKKOEHEELVQTVNTD	898
QY	1124	FSPYLSDNLEQPMGSSHASQVCSFETPDLLDDGEIKEDTSFAENDIKESSAVFSKSQVK	1183
Db	899	FSPCLISDNLEQPMGSSHASEVCSFETPDLLDDGEIKEDTSFAENDIKESSAVFSKSIQR	958
QY	1184	GELSRSPSPFTHLQAQYRGAKKLESEENLSBDEELPCFQHLLFGKVNIPQSOSTR	1243
Db	959	GELSRSPSPFTHLQAQYRGAKKLESEENLSBDEELPCFQHLLFGKVSNIQSOTTR	1018
QY	1244	HSTVATECLSKNTEENLSLKNLSINDCSNOVILAKASQEHHLSEETKCSASFSSQCSSEL	1303
Db	1019	HSTVATECLSKNTEENLSLKNLSINDCSNOVILAKASQEHHLSEETKCSGLFSSQCSSEL	1078
QY	1304	EDLTANTQDPPFLIGSSKOMRHSQSGVGLSDKELVSDDERGTGLENNQDEQSMDS	1363
Db	1079	EDLTANTQDPPFLIGSSKMRHSQSGVGLSDKELVSDDERGTGLENNQDEQSMDS	1138
QY	1364	NL 1365	
Db	1139	NL 1140	
RESULT 7			
O46488			
ID	O46488	PRELIMINARY;	PRT; 1146 AA.
AC	O46488;		
DT	01-JUN-1998 (trEMBLrel. 06, Created)		
DT	01-JUN-1998 (trEMBLrel. 06, Last sequence update)		
DT	01-JUN-2001 (trEMBLrel. 17, Last annotation update)		
DE	Breast cancer type 1 susceptibility protein (Breast and ovarian cancer		
DE	susceptibility protein) (Fragment).		
GN	BRCA1.		
OS	Alouatta seniculus (Howler monkey).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Alouattinae;		
OC	Alouatta.		

OX	NCBI_TaxID=9503;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=98122577; PubMed=9462745;		
RA	Hacia J.G., Makalowski W., Edgemon K., Erdos M.R., Robbins C.M.,		
RA	Fodor S.P.A., Brody L.C., Collins F.S.;		
RT	"Evolutionary sequence comparisons using high-density oligonucleotide		
RT	arrays.";		
RL	Nat. Genet. 18:155-158(1998).		
CC	-!- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. MAY PLAY A ROLE		
CC	IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).		
CC	-!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).		
DR	EMBL; AF019079; AAC39587.1; -.		
DR	InterPro; IPR002378; BrstCancer1.		
DR	PRINTS; PR00493; BRSTCANCER1.		
KW	DNA-binding; Nuclear protein; Anti-oncogene.		
FT	NON_TER 1 1		
FT	DOMAIN 278 285		
FT	DOMAIN 200 203		
FT	DOMAIN 815 818		
FT	DOMAIN 1001 1004		
FT	DOMAIN 996 1005		
FT	NON_TER 1146 1146		
SQ	SEQUENCE 1146 AA; 128119 MW; BFD07EBBF851890E CRC64;		
Query Match 52.0%; Score 5014.5; DB 6; Length 1146;			
Best Local Similarity 86.2%; Pred. No. 7.9e-251;			
Matches 991; Conservative 55; Mismatches 90; Indels 13; Gaps 7;			
QY	226	CEPSETDVTNTEHHQPSNNDLNTTEKRAERHPEKYGGSSVSNLHVPCGTNTHASSLOH	285
Db	2	CEPSEKDITNTEYHQSSNNDLNTTEKHATEKHPEKQCQSSVSNLHVPCGTNTHASSLOH	61
QY	286	ENSSLLLTDRMNVKAEFCNKSQPLARSOHNRWAGSKETCNDRTSTTEKKVYDLNAD	345
Db	62	ENSSLLLTDRMNVKAEFCNKSQPLARSOHNRWAESEETCNDRTSTTEKKVYDAD	121
QY	346	PLCERKENWKQKPCSENPR-DTEVDVWITLNSISIOKVNEWFSRDELGSDSDHGESE	404
Db	122	PLHGRKENWKQKPCSENPRDDTEDVAMLNSISIOKVNEWFSRDELTSDDSHDGGSE	181
QY	405	SNAKVADVLVNEVDYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKTFGKT	464
Db	182	SNAKVAEALVLEVDYSGSSEKIDLLASDPHDHLICKSERVHCKSVESIEDKTFGKT	241
QY	465	YRKASILPNLSHVTENLIICAFVTEPQIIQERPLTNKLRKRPTSLGHPEDFIKKADLA	524
Db	242	YRKASILPNLSHVTENLIICAFVTEPQIIQERPLTNKLRKRPTSLGHPEDFIKKADLA	301
QY	525	VQKTEMINOQGTNQTQNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKA	584
Db	302	VQKTEKINQGTNQTQNRNQVMNITNSGHENKTKGDSIQNENPNPVSLEKES-FKSKA	360
QY	585	EPISSSISNMELELNTHNSKAPKKNLRKRKSSSTRHHALELVVSRNLSPPNCTELQIDSC	644
Db	361	EPISSSISNMELELNTHNSKAPKKNLRKRKSSSTRHHALELVVSRNLSPPNCTEVIDSC	420
QY	645	SSSEETKKKKYNOMPVRHRSRNLMQMEGKEPATGAKSKNKPNEQTSKRHSDTTPPELKLTN	704
Db	421	SSSEETKKKKYNOMPVRHRSRNLMQMEGKEPATGAKSKNKPNEQTSKRHASDTTPPELKLTN	480
QY	705	APGSFTKCSNTSELKEFVNPSLPREEKEKLETVKVSNNNAEDPKDMLSGERVLOQERSV	764
Db	481	IPGSFTNCSTNEFKEFVNPSLPREEKEKLETVKLSNNAKDPKDLMLSGESVLOIERSV	540
QY	765	ESSISILVPGTDYGTQESISLLEVSSTLGAKTEPNKVCVSOCAAFENPKGLIHGCSKDNRN	824
Db	541	ESSISILVPGTDYGTQESISLLEVSSTLGAKTEPNKVCVSOCAAFENPKGLIHGCSKDNRN	600
QY	825	DTEGFKYPLGHEVNHHSRETSIEMEESELDQYLONTFKYVKRQSFALFSPNPGNAEECAT	884
Db	601	GTEGLKYPGPEVNYSOETSIDMRESELDTQYLONTFKYVKRQSFALFSPNPGNPEKCAT	660

QY 885 FSAHSGLLKKQSPKVTPECEQKEENOGKNESNIKPVQTVNITAGFPVVYQGDKPKVDNAKC 944
Db 661 FASCSRLAKQSPKVTPECEQKEENOGKEINELVETVNTTAGFPMVCQDKPKVDYARC 720
QY 945 SIKGSRFCLSQFRGNETGLITPNKHGLLQNPYRIPLPFKTSPVKTKCKNLLNEE 1004
Db 721 -IEGSRCLSLSQFRGNETGLITPNKHGLLQNPYHMSPLIPTRSFVKTKCKNLLNEE 779
QY 1005 EHSMSPEREMGNEN-IPSTVSTISRNNTRENVEFKASSNNI-----NEVGSSTNEVGS 1056
Db 780 EHSMSPERAMGNKIIPSTVSTISINN-RENAFKETSSSIYEVGSSTNEAGSSTNEVGS 838
QY 1057 SINEIGSSDENIQAELGRNRGPKLAMLRLGLVQIPEVYKSLPGSNCKHPKIKQYBEV 1116
Db 839 SINEVGSSENIQAELGRNRGPKLAMLRLGLVQIPEICKQSLPISDCKHPKIKQHEEV 898
QY 1117 VQTVNTDFSYLISNLEOPMGSSHASQVSETPDLDLDDGEIKEDTSAENDIKESAV 1176
Db 899 VQTVNTDVSCLISYLNLEOHMGSSHTSQVSETPDNLDDGEIKEDTSFAEYGIKETSTV 958
QY 1177 FKSQVKGELSPSPFTHTHLAQYRRGAKKLESSEENLSSEDELPFOHLLFGKVN 1236
Db 959 FKSQVKGELSPSPFTHTHLAQYRRGAKKLESSEENLSSEDELPFOHLLFGKVN 1018
QY 1237 IPSQSTRHSTVATECLSKNTNENLLSKNSLNDSCNQVILAKASOEHLSEETKCSASLF 1296
Db 1019 TP-QSTRHSTSAECLSKNTNENLLSKNSLSDSCNQVILAKASOEHYLSEETKCSGSLF 1077
QY 1297 SSQCSLEDLTANTYTQDPFLIGSKQMRHQSOSQVGLSKELVSDDEBRTGLERNQ 1356
Db 1078 SSQCSLEDLTANTYTQDPFLIGSKQMRHQSOSQVGLSKELVSDDEBRTGLERNQ 1137
QY 1357 EQGSDMSNL 1365
Db 1138 EQGSDMSNL 1146
RESULT 8
Q921D2 ID Q921D2 PRELIMINARY; PRT; 1812 AA.
AC Q921D2;
DT 01-MAY-1999 (TremBLrel. 10, Created)
DT 01-MAY-1999 (TremBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Brca1.
GN BRCA1 OR BRCA1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=97026289; PubMed=8872468;
RA Szabo C.I., Wagner L.A., Francisco L.V., Roach J.C., Argonza R.,
RA King M.C., Ostrander E.A.;
RT "Human, canine and murine BRCA1 genes: sequence comparison among
RT species."
RL Hum. Mol. Genet. 5:1289-1298(1996).
CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR EMBL; U68174; AAD00168.1; -;
DR MGD; MGI:104537; Brca1.
DR InterPro; IPR001357; Brct.
DR InterPro; IPR002378; Brst_cancerI.
DR InterPro; IPR001547; GH_5.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00533; BRCT; 2.
DR Pfam; PF00097; zf-C3HC4; 1.
DR PRINTS; PR00493; BRSTCANCERI.
DR SMART; SM00292; BRCT; 2.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00172; BRCT; 2.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; UNKNOWN_1.

DR PROSITE; PS00518; ZF_RING_1; 1.
KW Zinc-finger.
SQ SEQUENCE 1812 AA; 198803 MW; 1A61B60230365B63 CRC64;
Query Match 50.6%; Score 4885; DB 11; Length 1812;
Best Local Similarity 56.0%; Pred. No. 7.2e-244;
Matches 104; Conservative 265; Mismatches 485; Indels 70; Gaps 32;
QY 1 MDLSALRVEEVQNVINAMQKILECPICILELIKPEVSTKCDHIFCKFCMLKLLNKKGPSSQ 60
Db 1 MDLSAVQIQEVNVILHAMQKILECPICILELIKPEVSTKCDHIFCKFCMLKLLNKKGPSSQ 60
QY 61 CPLCKNDITKRSLOESTRFSQSLVEELLKIIICAFQDITGLEAYNSYNAFAKKNNSPEHLKD 120
Db 61 CPLCKNEITKRSLOGSTRFSQSLAEELLRIAAFEVDITGMQLTNGFSFSKKNNSCERLNE 120
QY 121 EYVSIQSGYNNRAKRLQSEPNPSLOETSLVSQNLGTVRLTQRKIQRIQKTSVYI 180
Db 121 EASITQSVGYNNRVRLPQVEPGNATLKD-SLGVQLSNLGTVRSYKKNRQTQPRKKSYYI 179
QY 181 ELGSDSSEDTVNKATYCSVGDOELLQITPQGTDRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 180 ELSDSSSEETVKPGDCSVRODELLQTAPQAGDEGKLHSAEEAAEFSE-GIRNIEHHQ 238
QY 241 PSNNDLNTTEKRAAERHPEKYQGSSVSNLHVPCGTNTHASSLQHENSSLTLTKDRMNV 300
Db 239 CS-DLNPNTENHATERHPEKQCSISINVCVPCGTDAHASSLOPETSLLLIEDRNAE 297
QY 301 KAEFCNKSQKQGLARSHQNRWAGSKETCNDRTPTSTKRYDLNADPLCERKWNKOKLPC 360
Db 298 KAEFCNKSQKQGLARSHQNRWAGSKETCNDRTPTSTKRYDLNADPLCERKWNKOKLPC 357
QY 361 SENPRDTEVPWITLNSIOKVNWFSDSLGSDSDHSDGESNAKVADVLDVLEVD 420
Db 358 PENSAGTDDVPWITLNSVSKVNFWSRTGEMLTSDSASARRHESNAAVAVLEVSNEVD 417
QY 421 EYSGSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKYRKKASPLNLSHVTEN 480
Db 418 GGFSSSRKTLDTVPDHPHTLACKSGRDFSKPVEDNISDKIFGKSVQRKSGRPHLNHYTE- 476
QY 481 LIICAFVTEPOIIOERPLTNKLRKRPSTGLHPEDFTKKADLA-VQKTPEMINGOTNQT 539
Db 477 -IITFTITEPOITQEQPFTNKLKRK--STSLQPEDFTKKADSAQVQRTPDNINOQTDLM 533
QY 540 EQNGVWNITNSGHENKTGDSIONENPNPDESLEKESAEKTAEPISSTSSNNELELN 599
Db 534 EPNEQAVSTTSNQENQIAGSNLQKESAPTESLRKEPASTAGAKSISNSVSDLEVLN 593
QY 600 IHNSKAPKNNLRKSSSTRHITHALELVVSRNLSPNCTVELQIDSCSSSEETKKNYNQMP 659
Db 594 VHSSKAPKNNLRKSSSTRHITHALELVVSRNLSPNCTVELQIDSCSSSEETKKNYNQMP 652
QY 660 VRHSRNLQMEGKEPATGCAKSNKPNQETSKRHSDDTPELKLTNAPGSFTKCSWTSELK 719
Db 653 AGHLREPLQIEDTEPAADAKK-NEPNEHIRKRRASDAPPEEKLMNKAAGLLTSCSPRSQ 711
QY 720 EFVNPSLPREEKEKLETVKYSNNAEDPKDLMLSGE-RVLOTERSVESSTSLVPGTDVG 778
Db 712 GPNVPS-PORTGTEQLTQMSDSAKELGDRVLGGEPSGKTTDRSEESTSVSLVPTDITD 770
QY 779 TQESISLLEVTGLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNNDNTEGPKYPLGHEVN 838
Db 771 TQNSVSVLDAHTVRYARTGSAQCMTQFVASENPKELVHG-SNNAGSGTEGLKPLRLHALN 829
QY 839 HSRTSITEMESELDAQYLQNTFKVSRQSFALSPNPGNAEEBECATFSAHSGSLKKQSPK 898
Db 830 LSQE-KVEMEDSELDQYLTQNTFQVSRQSFALSPKPSQKDC---AHSVPSKELSPK 884
QY 899 VTFPECKEENOGKNESNIKPVQTVNITAGFPVVYQGDKPKVDNAKCSIKGSRFCLSSQF 958
Db 885 VTAQKQK-ERQGOEEFEISHVQVAATVGLPVLICQEGKLAADTWCD--RGRSLCPSSH 941
QY 959 RGNETGLITPNKHGLLQNPYRIPLPFKTSPVKTKCKNLLNEENFEHSMSPEREMGNEN 1018

```
Db 942 RSGNGLSATGKSGISONSFHFKQSVSPIRRSIKTDNRKPLTEGFRFHTSTSTEMAVGNEN 1001
QY 1019 I-PSTVSTISNNIRENVFKEASSNINEVGSSSTNEVGSSDENEIQAEFLGRNRG 1077
Db 1002 ILQSTVHTVSLNN-RGNACGAGS-----GSIHEVCSTGDSFGPGLGRNRG 1046
QY 1078 PKLNAMLRLGVLOPEVYKQSLGNSCKHPEIKKQYEEVVQTVNTDFSPYLLISDNLBQPM 1137
Db 1047 PKVNTVPLDSMQPGVQCQSVPSD-KYLEIKKQGEAVC----ADFSPLCFSDHLQSQM 1101
QY 1138 GSSHASQVCSETPDLLDDGEIKEDTSFAENDIKESSAVFSKSVQKGLSRSPSPFTHTH 1197
Db 1102 -SGKFVQVCSETPDLLDDVEIOGHTSFGEGDIMERSAVFNGSILRRRESSRSPSPVTHAS 1160
QY 1198 LAQGYRGAKKLESSENLSSEDELPFOHLLFGKVNNTIPSQSTRHSTVATECLSKNTE 1257
Db 1161 KSQLHRASKRLSESSEDSSTEDDLPCFOHLL-SRLSNTP-ELTRCAVSATGOIPEKAE 1218
QY 1258 ENLLSKNSLNDCSNQVILAKAQEHLSSEETKCSASLFSQSCSELEDLTANTNTODPFL 1317
Db 1219 GTQAPWKSSSDCNEVIMTEAQEHQFSEDPRCSGRMFSSQNSAAQGSTANANSQDSNF 1278
QY 1318 IGSSKQMRHOSGCGVGLSKDELVDDEERGTLGLENQNEEQSDMSNLGEA-ASGCSEST 1376
Db 1279 IPPSKQRSHQCGNEAEFLSKELISDNEEMATCLEEDNDQEE--DSIIPDSEASGYESET 1336
QY 1377 SVSEDCSGLSSQSDILTTQORDTMOHLNLIKLOQEMAELEAVLEQHGQSPNSYPSIISDS 1436
Db 1337 NLSEDC-----SQSDILTTQORDTMOHLNLIKLOQEMAELEAVLEQHGQSPNSYPSIISDS 1392
QY 1437 SALEDLRNPQSTSEKAVLTQSOKSEYPISONPEGLSADKFEVS-ADSSTSKNKEPGVER 1495
Db 1393 CALEDLPDLEPNKSGAAILTSKNINENPVSONLKSACDDKRFQLQHLGEGPTSGDDESGMR 1452
QY 1496 SSPSKCPSLDRWYHSCSGSLQNRNYPQBELIKVYVDVEEQLEESGPHDLTETSYLPR 1555
Db 1453 PSPPKSLAGRSGAHGCSRHLQKRNSPQBELLPQAGSE-----ASSEPHNSTGQSCCLPR 1508
QY 1556 QDLEGTPLESGLSFLSD-DPESDPSDRAPESARVGNIPSPSTSAKVPOPKVAESAQSP 1614
Db 1509 RELEGTPYLSGSLFSRDPSESPK-----EPAHIGTTTASTSALKIPOGQVAFRAAAA 1564
QY 1615 AAHTTDTAGYNAMEESVREKPELTASTERVNKRMSVMYSGLTPEEFMLVYKFAKHHI 1674
Db 1565 AGAD-----KAVVGIVSVKIKPELTSEERADRIDSMVYSGLTPEEFMLVYKFAKYRL 1617
QY 1675 TLNLTITEETHVVMKTDAEFVCERTILKYFLGTAGGKVVYVFWYQTSIKERKMLNEHDF 1734
Db 1618 TLTDAITEETHVVIITDAEFVCERTILKYFLGTAGGKVIYVSWVYRSIQERRLLNVHEF 1677
QY 1735 EVRGDVYNGRNGHGPKRARESODRKIFRGLGICCYGPFNTNMPDOLBMMVQLCGASVYKE 1794
Db 1678 EVKGDVYNGRNGHGPKRRESRE-KLFGKLVYCCPEFTNMPKDELERMQLCGASVYKE 1736
QY 1795 LSSFTLGTGVHPVIVVQPDADWEDNGFHAIGQMCPEAPVYVRENVLDVALYQCELDLYL 1854
Db 1737 LPSLTHTDGAHLVIVQPSAWTEDSNCPDGLQCKARLVNMDVLDLSLSSYRCRDLDAYL 1796
QY 1855 IPQI 1858
Db 1797 VQNI 1800
```

RESULT 9
O54952

```
ID O54952 PRELIMINARY; PRT: 1817 AA.
AC O54952; p97951;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Breast cancer type 1 susceptibility protein (Breast and ovarian cancer
susceptibility protein).
```

```
GN BRCA1.
OC Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=99111388; PubMed=9892727;
RA Bennett L.M., Brownlee H.A., Hagavik S., Wiseman R.W.;
RT "Sequence analysis of the rat brca1 homolog and its promoter region.";
RL Mamm. Genome 10:19-25(1999).
RN [2]
RP SEQUENCE OF 8-222 FROM N.A.
RC STRAIN=WISTAR-KYOTO; TISSUE=SPLEEN;
RX MEDLINE=96358532; PubMed=8761410;
RA Chen K.S., Shepel L.A., Haag J.D., Heil G.M., Gould M.N.;
RT "Cloning, genetic mapping and expression studies of the rat Brca1
gene.";
RL Carcinogenesis 17:1561-1566(1996).
CC -!- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION, MAY PLAY A ROLE
IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
DR EMBL; AF036760; AAC36493.1; -.
DR EMBL; S82504; -; NOT_ANNOTATED_CDS.
DR EMBL; S82502; -; NOT_ANNOTATED_CDS.
DR EMBL; U60523; AAB40387.1; -.
DR EMBL; S82500; AAB37501.1; -.
DR InterPro; IPR001357; BRCT.
DR InterPro; IPR002378; Brst_cancer1.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00533; BRCT; 2.
DR Pfam; PF00097; zf-C3HC4; 1.
DR PRINTS; PR00493; BRSTCANCER1.
DR SMART; SM00292; BRCT; 2.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00518; ZF_RING_1; 1.
DR PROSITE; PS00518; ZF_RING_1; 1.
KW Zinc-finger; DNA-binding; Nuclear protein; Anti-oncogene.
FT ZN_FING 24 64 C3HC4-TYPE.
FT DOMAIN 497 503 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT CONFLICT 38 38 Q -> K (IN REF. 2).
FT CONFLICT 192 192 A -> M (IN REF. 2).
SQ SEQUENCE 1817 AA; 199876 MW; C0B4760F0E349A01 CRC64;
```

Query Match 50.5%; Score 4873.5; DB 11; Length 1817;
Best Local Similarity 56.2%; Pred. No. 2.9e-243;
Matches 1049; Conservative 259; Mismatches 488; Indels 69; Gaps 32;

```
QY 1 MDLSARVEEVQVNVINAMQKILECPICLELIKBPVSTKCDHIFCKMCLKLNQKGPQ 60
Db 1 MDLSARVIEQVNVIAHQKILECPICLELIKBPVSTQCDHIFCKMCLKLNQKGPQ 60
QY 61 CPLCKNDITKRSQESTRFSQVLEELLKIICAFQDITGLYANSYNAFKKNSPEHLKD 120
Db 61 CPLCKNEITKRSQESARFSQVLEELLKIIDAIFELDTGMOCANGSFSEKKNSSSELLNE 120
QY 121 EVSIIOQMGVNRKAKRLQLQSEPNPSLQETSLSVQLSNLGTVRTLTQRIOQKTSVYI 180
Db 121 DASIIQSVGYNRVKKLQIQESGATLKD-SLSVQLSNLGIIVSRMKNKRNQTPONKSVYI 179
QY 181 ELGSDSSEDTVNKATYCSVGQDQLLQITPOGTRDEISLDSAKKAACEFSQDVTNTEHQ 240
Db 180 ALESDSSEERVNAPDGCVSVDQLFQIAPGAGDEKLSAKKAACDFSE-GIRNIEHQ 238
QY 241 PSNNDLNTEKRAAERHPEKYQGSVSNLHVEPCGNTNTHASSLQHENSSLLLTKDRMNV 300
Db 239 CSDKDLNPTENHATERHPEKCPISVANVHVEPCGTDARASSLQRTGRTSLTLFTEDRLDAE 298
QY 301 KAEFCNKSQOPGLARSOHNRWAGSKETCNDRRPTSTPEKKYVDLADPLCEKRNKQKLP 360
Db 299 KAEFCDRSKSGAAVSQQRWADSKETCNGRPVPRTEGKADPNVDSLGRKNQHNPKSLC 358
```

```
QY 361 SENPRDTEVPWITLNSSIQKYNWFPSRDELIGSDSDHGESESNKAVYADVLDVNEVD 420
  || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 359 PENSGATTDPWITLNSSIQKYNWFPSRGTGEMLTSDNASDRPASNAEAAVLEVSNEVD 418
  || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDIKFGKTYRKKASPLNLSHVTE 480
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 419 GCFSSKKIDLAVDPDNVAMTSGTRDFSKPVENIINDIKFGKTYRKGRSPLNHNVE- 477
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 481 LIICAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFTKKADL-AVQKTPEMINGOTWOT 539
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 478 -IIGTFTEPQIIQEPFTNKLKRR--STCLHPEDFTKKADLTYYQRISENLNQTDQM 534
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 540 EONGQVYNTNSGHENKTKGDSIQENKPNPTIESLEKESAPTKAEPTSSISNMELELN 599
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 535 EPDQAMSTISNGQENRATGNDLQRCRNAPHPIESLRKEPAFTAKAKSISNSISDLEVELN 594
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 600 IHNSKAPKNRLRRKSSRIHIALELVLSRNLSPPNCTELQIDSCSSSEETKKKYNOMP 659
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 595 VHSKAPKNRLRRK-STRCVLPLE-PISRNPSPPTCAELQIESCGSSEETKKNNSQTP 652
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 660 VRHSRNLQMECKEPATGAKKSNKNEQTSKRHSDTTPPELKLITNAPGSFYKCSNTSELK 719
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 653 AGHIREPQIUEDTEPAADAKK-NEPNEHIRKRSASDAFPEEKLMMKAGLLTSCSPRPKP 711
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 720 EFVNPSPREEKEKLETVKVSNNAEADPKDMLSGE-RVLOTPERSVESSSISLVPQTDYG 778
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 712 GPVNPSPERKIGIEQLEMCOMPNNKELGDLVLGEPGSKPTEPSEESTSVSLVPDQTDYD 770
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 779 TQBSISLLEVSTLGAKTENPKVCVQCAFENPKGLIHGCKDNDRNDETFKYPILGHEVN 838
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 771 TQNSVILEANTVRVARTGSCVQMTQFVASENPKNELVHG-SNNAGSGSECFKHLRHELN 829
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 839 HSRETSIEMESELDAQYLQNTFKVSKQSPALFSPNPGNAEEBCATFSAHSGSLKKQSPK 898
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 830 HNOET-IEMEDSELDTQYLQNTFQVSKQSFALFSLKRSPOKDTLVGARSVPSPREPSPK 888
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 899 VFPECKEENOGKNEKNIPVOTNITAGPPVGVGOKPVDNNAKCSIKGGSRFCLSQF 958
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 889 VTSRGQK-ERQGESELSHVQAVTVVGLVPVQCEKPGAVTMCA--DVSRLCPSSH 945
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 959 RGNETGLITPNKHGLQNPYRPIPLPIKSFVKTKCKKNLLEENFEHSMSPEREMNGEN 1018
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 946 RSCENGLNTDKSGISQNSHFQSVSPLSRSIKTDNRKLTTEGREKUT---ERGMGNET 1002
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1019 -IPSTVSTISRNIENYFKEASSNINEVGSSTNEVGSSINEIGSSDENTQOALGRNRG 1077
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1003 AVOSTIHTISLNN-RGDACLEASS-----GSVIEVHSTGENYOGOLDNRG 1047
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1078 PKLNAMLRGLVLPQVEYKQSLPGSNCKHPKIKQYEEVWQTVNTDFPYLISONLEQPM 1137
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1048 PKVNTVSLDSTQPGVSKQSAVSD-KYLEIKQES-----KAVSADFSPCLFSDHLEKPM 1101
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1138 GSHASQVCSPTDLDLDDGETIKEDTSPAENDIKESSAVFSKSVQKGBELSRSPSFTHH 1197
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1102 RSDKTFQVCSEPTDLDLDDVETQENASGEGGITEKSAIFNGSVLRRSSRSPSPVTHAS 1161
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1198 LAGGYRRGAKKLESSENLSSDEELPCFOHLLFGKVNNIPQSOTRHSHTVATECLSKNTE 1257
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1162 KRSULHRGSRKLEFSEESDSTDEDLPCFOHLL-SRVSSSTP-ELTRCSSVYTVRPEKAK 1219
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1258 ENLLSLKSLNDCSNQVILAKASQPHLSSEETKCSASFSSQCSLEDLTANTQPPFL 1317
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1220 GTQAPRKSSIDCNNEVILGEASQEQYSEDAKCSGSMFSSQHSAAALGSPANALSQPDF 1279
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1318 IGSSQMRHQSSQGVGLSDKELVSDDEERTGLLEE-NNQEQSM--DSNLGEAASGES 1374
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1280 NPPSKQRHQAEENEAFLSKDELISDHEDMAACLEAASQDEEDSIIPDS-----VASGVES 1335
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1375 ETVSVSDCSGLSSQSDILTTQORDTQMHLNKLQOEAMAELEAVLEOHGSQPSNSVPSTIS 1434
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1336 EANLSEDC-----SQSDILTTQORDTQMHLNKLQOEAMAELEAVLEOHGSQPSGHPPLCPA 1391
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 1435 DSSALEDLRNPQSTSEKAVLTSSQSSSEYPISONPEGLSADKFEYS-ADSSSTSKNKEPGV 1493
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1392 DPCALEDLDPDEQNRSGTALTTSKNINENPVSONPKRACDDKASQPPDGLSPGDKESGM 1451
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1494 ERSSSKCPSLDDRWYMHSCSGSLQNRNYPSOEELIKVVYVEEQOLESQPHDLTETSYL 1553
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1452 RPSPFKSPLTSSRCSARGHSRSLQNRNSTSQEELLQPAKLE---KSCPEHNLTRGSL 1507
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1554 PRODLEGTPLYESGLSLESDPDESDEAPESARVGNIPSSSTALKVPOLKVAESAQS 1613
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1508 PRODLEGTPLYESGLSLESDPDESDEAPESARVGNIPSSSTALKVPOLKVAESAQS 1564
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1614 PAAAHHTDTAGYNAMESVREKPELTASTERNVRKMSMVVSGLTPEEBMLVYKPKAKHH 1673
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1565 PAAG-GADT---AVVEIVSKIKPEVTPSKERAERDISMVVSGLTPEKVIYQKFAEKYR 1619
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1674 ITLNLITEETHVVMKTDABFVCERTLYKFLGAGGKMWVSFYVWTSQIKERKMLNEHD 1733
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1620 LALTDVITEETHVVIKTDAEFVCEERTLYKFLGAGGKMWVSFYVWTSQIKERKMLNEHD 1679
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1734 FEVRGDVYNGRNHOGPKRARESODRKIFRGLEICCYGPTNNMPTDOLEHMMVOLCGASVVK 1793
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1680 FEVRGDVYNGRNHOGPKRARESODRKIFRGLEICCYGPTNNMPTDOLEHMMVOLCGASVVK 1738
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1794 ELSSTFTLGTGVHPIVVVQPDWATDNGFHAIGOMCEAPVVTREWVLDSVALYQCOBELDY 1853
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1739 ELPLLTRDTGAHPVLVQPSAWTEDNDCPDIGQLCKGRVMDWDLDSISVYRCRDLDAY 1798
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1854 LIPQI 1858
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1799 LVQNI 1803
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
O46490 PRELIMINARY; PRT: 1146 AA.
AC O46490;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Breast cancer type 1 susceptibility protein (Breast and ovarian cancer
DE susceptibility protein) (Fragment).
GN BRCA1.
OS Cynoccephalus variegatus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Dermoptera; Cynocephalidae; Cynoccephalus.
OX NCBI_TaxID=9457;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98122577; PubMed=9462745;
RA Hacia J.G., Makalowski W., Edgemon K., Erdos M.R., Robbins C.M.,
RA Fodor S.P.A., Brody L.C., Collins F.S.;
RT "Evolutionary sequence comparisons using high-density oligonucleotide
RT arrays.";
RL Nat. Genet. 18:155-158(1998).
CC -!- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. MAY PLAY A ROLE
CC -!- IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
DR EMBL: AF019081; AAC39589.1; -.
DR InterPro: IPR002378; Brst_cancer1.
DR PRINTS: PR00493; BRSTCANCER1.
KW DNA-binding; Nuclear protein; Anti-oncogene.
FT NON_TER 1
FT DOMAIN 276 283 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 818 821 POLY-SER.
FT NON_TER 1146 1146
SQ SEQUENCE 1146 AA; 128330 MW; FD23388B602380AE CRC64;

Query Match 45.5%; Score 4386.5; DB 6; Length 1146;
Best Local Similarity 76.2%; Pred No. 2.2e-218;
Matches 874; Conservative 100; Mismatches 166; Indels 7; Gaps 6;

QY 225 ACEFSETDVTNTEHHQPSNNDLNTEKRAERHPKEKYOGSSVSNLHVPCGTNTHASSLQ 284
```

Db 1080 AVEDLATNTNTQDPFLMFDPSPNHNVRHQSENVVLSDKRELVSDDERETALEYNHEEE 1139

QY 1359 QSWDSNL 1365

Db 1140 QSVDSNL 1146

RESULT 11

O46489 PRELIMINARY; PRT; 1141 AA.

AC O46489;

DT 01-JUN-1998 (TRENBLrel. 06, Created)

DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)

DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)

DE Breast cancer type 1 susceptibility protein (Breast and ovarian cancer susceptibility protein) (Fragment).

GN BRCA1.

OS Galago crassicaudatus (Thick-tailed galago) (Otolemur crassicaudatus).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Strepsirhini; Galagonidae; Otolemur.

OX NCBI_TaxID=9463;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98122577; PubMed=9462745;

RA Hacia J.G., Makalowski W., Edgemon K., Erdos M.R., Robbins C.M., Fodor S.P.A., Brody L.C., Collins F.S.; "Evolutionary sequence comparisons using high-density oligonucleotide arrays.";

RT Nat. Genet. 18:155-158(1998).

RL FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. MAY PLAY A ROLE IN CELL PROLIFERATION AND DIFFERENTIATION (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).

DR EMBL; AF019080; AAC39588.1; -.

DR Interpro; IPR002378; Brst_cancer1.

DR PRINTS; PRO0493; BRSTCANCER1.

KW DNA-binding; Nuclear protein; Anti-oncogene.

FT NON_TER 1

FT DOMAIN 277 284 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).

FT DOMAIN 199 202 POLY-SER.

FT DOMAIN 816 819 POLY-SER.

FT NON_TER 1141 1141

SQ SEQUENCE 1141 AA; 127578 MW; 2653CA894834946F CRC64;

Query Match 41.9%; Score 4040; DB 6; Length 1141;

Best Local Similarity 72.1%; Pred. No. 1.7e-200;

Matches 831; Conservative 105; Mismatches 192; Indels 24; Gaps 15;

QY 226 CBFSETDVTNTEHHQPSNNDLNTEKRAAEHPKPYQGSVSNLHVPCGTNTHASSLQH 285

Db 2 CMVSEKDIITSIEHQSSNKDLNTEKRATEMHPKPYQGSVSNLHVPCGKWNTHASSLQH 61

QY 286 ENSLLLTQDRNVKAEFCNKSQKPLARSQHNWAGSKETCNDRRTPTSEKKVDLND 345

Db 62 ESSLLLTQDRNVKAEFCNKSQKPLARSQHNWAGSKETCNDRHTCSPKQKVDLNTA 121

QY 346 PLCKERKWKQKLPSCENPRDTEPWITLNSIQKVNWFSDLLGSDSDSHGSES 405

Db 122 PYGRKEQKELKLLCSKNRPSODVPWITLNSIQKVNWFSDSDSHGSES 181

QY 406 NAKVADVLVNEVDYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTY 465

Db 182 HAEVAGALEVPEDVGYSSSKIDLLASDPHYPIICKSERVHSKPIKSKVEDKIFGKTY 241

QY 466 RKKASLPNLSHVTEENLIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPEDFIKKADLA - 524

Db 242 RKKASLPNLSHVTEENLIIRAAATEPQITQECSLTNKLRKRPTSGLCPEDFIKKADLAV 301

QY 525 VOKTPEMINQGTNQTQONQOVNITNSHENTKGDSONEKNPNPIESLEKSAFKTKA 584

Db 302 VOKTPEKRIQGTNQNVDONSHVNVITNSGYNETKGDYQVONEKNANSTESLEKESLGTKA 361

QY 585 EPTSSISNMELELNTHNSKAPKKNRLRKRSTRHHALELVVSRNLSPNCTELQIDSC 644

Db 1 ACESFSE-DITNIEHQSCDKDLNTEKHATERHPKPYQSSVSNLHVPCGTNTHASSLQ 59

QY 285 HENSLLLTQDRNVKAEFCNKSQKPLARSQHNWAGSKETCNDRRTPTSEKKVDLNA 344

Db 60 HENSLLLTQDRINVEKTEFCNKSQKPLARSQHNWAGSKETCNDRPTSTEKKIDLNA 119

QY 345 DPLCKERKWKQKLPSCENPRDTEPWITLNSIQKVNWFSDLLGSDSDSHGSE 404

Db 120 DSQGRKERKWKQKLPSPEDPQDPWITLNSIQKVNWFSDLLGSDSDSRVSE 179

QY 405 SNAKVADVLVNEVDYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKT 464

Db 180 SNAKVAGALEVPEDVGYSSSKIDLLASDPHYPIICKSERVHSKPIKSKVEDKIFGKT 239

QY 465 YRKASLPNLSHVTEENLIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPEDFIKKADLA 524

Db 240 YRKASLPNLSHVTEENLIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPEDFIKKADLA 299

QY 525 -VOKTPEMINQGTNQTQONQOVNITNSHENTKGDSONEKNPNPIESLEKSAFKTK 583

Db 300 VOKTPEKINOQIDQEQNDRVNIIINSGHENETKDDYVQKEKNANPTESLEKSAFRK 359

QY 584 AEPISISSNMELELNTHNSKAPKKNRLRKRSTRHHALELVVSRNLSPNCTELQIDS 643

Db 360 AEPISISSNMELELNTHNSKAPKKNRLRKRSTRHHALELVVSRNLSPNCTELQIDS 419

QY 644 CSSESBEIKKKYNQMPVRISRLQMEGEPATGAKSNKPNQOTSKRHDSDTFPELKL 703

Db 420 CSSESBEIKKKYNQMPVRISRLQMEGEPATGAKSNKPNQOTSKRHDSDTFPELKL 479

QY 704 NAFGSFTKCSNTSELKEFNPSLPREEKEKLETVKVSNAEDPKDLMLSGERVLQTE 763

Db 480 NVPVFANCSNNKLEQFDPSLQREIEENLETIHVSNAKDPKDLVLSGEGKLGTE 539

QY 764 VESSISLVPGTDYGTQESISLLEVTLGAKTEPNKVCYQCAAFENPKGLIHGCSKDN 823

Db 540 VESTSISLVPGTDYGTQDSISILEANILGAKTAPSOHANQCAAIENPKELIHGCPK 599

QY 824 NDEGKPYPLGHVNHUSRETSIEMSELDQALQNTFKVSRQSFALSPNCGNAEECA 883

Db 600 NDEGKPYPLGHVNHUSRETSIEMSELDQALQNTFKVSRQSFALSPNCGNAEECA 659

QY 884 TFAHSGSLKQSPKVTFCQEQEENQKNEISNPKVQVNTITAGPVPVQK-DKPVDNA 942

Db 660 TVYHSELRLKQSPKVTFCQEQEENQKNEISNPKVQVNTITAGPVPVQK-DKPV 719

QY 943 KCSIKGSRFLCSSQFRGNETGLITPNKHGLLQNPYRIPPLFPKISFKVTKCKNLL 1002

Db 720 KFSIKGSRFLCSSQFRGNETGLITPNKHGLLQNPYRIPPLFPKISFKVTKCKNLL 779

QY 1003 FEHSHSPERMCNEN-IPSTVSTISRNIRNVFKEASSNINEVGSSTNEVGSSINEI 1061

Db 780 FEHSHSPERMCNEN-IPSTVSTISRNIRNVFKEASSNINEVGSSTNEVGSSINEI 839

QY 1062 GSSDENIQALGNRGPKNLNLRLGVLOPEVYKQSLPCSNCKHPKIKKQEEVEVQTVN 1121

Db 840 GPGENIQALGNRGPKNLNLRLGVLOPEVYKQSLPCSNCKHPKIKKQEEVEVQTVN 899

QY 1122 TDFSPYLIDNLEQPMGSSHASQVCSETPDDLLDGDGEIKEDTSFAENDIKESSAVFSKV 1181

Db 900 TNLCLSLHNLQPMGSSHASQVCSETPDDLLDGDGEIKEDTSFAENDIKESSAVFSKV 959

QY 1182 QKGLSRSPSPFTHLAQYRGAKKLESSENLSEDEELPCFOHLLFGKVNTPPSQS 1241

Db 960 QKGLSRSPSPFTHLAQYRGAKKLESSENLSEDEELPCFOHLLFGKVNTPPSQS 1019

QY 1242 TRHSTVATCLSKNTEENLLSKNSLNDSCNOVILAKASOEHLSEETKCSASLFSQS 1301

Db 1020 TRHSTVATCLSKNTEENLLSKNSLNDSCNOVILAKASOEHLSEETKCSASLFSQS 1079

QY 1302 ELBDLPANTNTQDPFLI--GSSQKMRHQSESQGVGLSKRELVSDDERETGLEE-NNQEE 1358


```
Db 362 EPISSISNMKLELNTHNSKASKKRLRKKSSRHIRALELVVNNPSPPHNTHLIQIDSC 421
QY 645 SSSEELKKKYNOMPVRHSRNLIQMEGEPATGAKKSNKPNEBOTSKRHSDSTFFPELKLTN 704
Db 422 SSSEELKDKSSDOIPIVHRSRKLCLMEDREPATGAKKSNKPNEQISKRHVSDFPEVALTN 481
QY 705 APOGFTKCSNTSELKEFVNPSLPREKEKL-ETVKVSNNAEDPKDLMLSGERVLIQ--T 760
Db 482 ISSFFTNCSGSNRLKEFVNPSLQKRTENLEETIQVSNSTKGP---VLSGERVLIQISE 538
QY 761 ERSVSSSLVPGTDYCTQESISLLEVSTLQKAKTEPNKCVSQCAAFENPKGLIHGCSK 820
Db 539 ERSIKTSVDPDITDYGTDNSLKVLRKVTPANKHASQGTATENKELIHGCSK 598
QY 821 DNNDTEGFKYPLGHEVNSHRSSTSEMESELDQYLONTFKVSKRQSFALFSPGNAAE 880
Db 599 DTGNDTEGKYDPLRHEINHIQISMEDESELDQYLONTFKFSKRQSFALFSLNIG---K 655
QY 881 ECATFSAH---SGSLKKQSPKVTFFCEQKEENOGKNESNIKPVQTNITAGPPVVGQDKP 938
Db 656 ECATVCAQSLSASLRKGS-KVILECEQ-IENPGMKPEKIKHIQGNINITGFSVVCQDKK 713
QY 939 VDN-AKCSITKSGSRCLSSQFRGNETGLITPNKHGGLLQNPYRIPPLPIKSVFTKCKKN 997
Db 714 TDYAKYSIKEASRFLCSNQFRNETESITVKNLGILOQLYHIPPLSPIRLFDKTKCNTN 773
QY 998 LLEENFEHSMSPEREMGNEN-IPSTVSTISRNNIRENVFEKASSSNINEVSGSTNEVGS 1056
Db 774 LLEEREHSHVLPKAVGNENTVPMTMTINQNN-RESAYKEASSSNINEVSSSTNEVGS 832
QY 1057 SINEIGSSDENIOAELGRNRGPKLAMLRLGVLOPEVYKQSLPGSNCKHPKIKQOEYEV 1116
Db 833 SYNEVGPSSENIOAELDKNRGPKLAVLRLGLMQPEVYKQNLPISNCEHPKIKQOENG 892
QY 1117 VQTVNTDFSPYILSONLEOPMGSSHASQVCSSETPDDLLDGEIKEDYSFAENDIKESSAV 1176
Db 893 VQVPNPDFSSCLISLONLEOPTRSSHASQVCSSETPDDLLVDDELKENTSFADNNIKERSAV 952
QY 1177 FKSVOGGLSRSPSPFTHLAQCYRRGAKKLESSEENLSSEDEELPCFOHLLFGKVNN 1236
Db 953 FSKNVMREISRSPSLAHILHTAQHREVRKLESSEENMSSEDEELPCFOHLLFGKVTN 1012
QY 1237 IPSQTRHSTVATECLSKNTENLNSLNSLNDSCNQVILAKASQEHHLSETKCSASLF 1296
Db 1013 TSQSSTQSTIATECLSRKTEETLVSQNTLNDSCNQVTLVKASQENHLSEAKCSVSFL 1072
QY 1297 SSQCSLEDLTANTTQDPFLIGS--SKOMRHQESQCVGLSKDELVSDDERGTGLEE- 1353
Db 1073 SSQCS--EDLTANTTQDPFLMSDPLSKOMRDQSENQEV-LSDNELVSNDEEREPSLEED 1129
QY 1354 NQOEQSDMSNL 1365
Db 1130 NQOEQSVSDSL 1141
RESULT 12
Q924E0 PRELIMINARY; PRT; 1063 AA.
AC Q924E0.
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BRCA1 (Fragment).
GN BRCA1.
OS Pedetes capensis (Springhaas).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Pedetidae; Pedetes.
OX NCBI_TaxID=10023;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21219190; PubMed=11319262;
RA Adkins R.M., Gelke E.L., Rowe D., Honeycutt R.L.;
RT "Molecular phylogeny and divergence time estimates for major rodent
```

```
RT groups: evidence from multiple genes.;
RL Mol. Biol. Evol. 18:777-791(2001).
DR EMBL; AF332047; AAK71627.1; -.
FT NON_TER 1
FT NON_TER 1063
SQ SEQUENCE 1063 AA; 118726 MW; 9B9CE758FE40745 CRC64;

Query Match      38.3%   Score 3693;   DB 11;   Length 1063;
Best Local Similarity 69.5%   Pred. No. 1.3e-182;
Matches 752;   Conservative 103;   Mismatches 203;   Indels 24;   Gaps 7;

QY 262 QGSVSNLHVPCGTTNTHASSLQSHENSLLLTDKRMNVKEAFECNCKSKOPGLARSQHNRW 321
Db 1 QGISVSNLCVPCGTTNTHASLLQHGSSNLLTKDMNVKEADFCNCKSKOPGLARSQHNRW 60
QY 322 AGSETCNDRTPTSTKVKYLDNADPLCERKEWNKQKLPCCSENPRDTEYDPWITLNSSTQK 381
Db 61 AESKETCNDRTPTSTKVKYLDNADPLFYGRKERKNQKPPCCSENPRDQVSWYTLNSSTQK 120
QY 382 VNEFWSRDELGLSDSDHGDSESENAKVADVLVNEVDEYSGSSEKIDLLASDPHEALI 441
Db 121 VNEFWSRDEMLTSDDLHDGVSESENAEVLGALEVPDEIKHGSGSEKIDLLASDPHNALT 180
QY 442 CKSERVHSKSVESNIEDKIFGKYRKASPLNSHVTENLIIGAFVTEPQIQIERPLTNK 501
Db 181 CKSERVCAKPVESNIEDKVFGRYRKISLPLNSPITENLVGAFATKPIQRQEHPLTNK 240
QY 502 LKRRRTSLGHPDFTKKADLAV-OKTPMINOGTNOQTEONGQVMNITNSGHEKNTKGD 560
Db 241 LKRRKATSFELRPEDFTKKVDLTVIQTPKINQTAQMGQGVQVNNNTNLENETKND 300
QY 561 SIQEKAPNPITSELESKESAFKTKAEPITSSISNMELENIHNSKAPKNRLRRKSSSTRI 620
Db 301 YVLKEKAPNPITSELESKESSTHRTKAEPISISNMELENIHNSKAPKNRLRRKSSSTRI 360
QY 621 HALELVYSRLSPNCTELQIDSCSSSEETIKKKYNQMPVHRHNSRLQMEGKEPATGAKK 680
Db 361 HALELVYNNKPSNPQTELQIDSCSSSEETIKKKCSQMSVRSRKLQLMEDAEPAAGAKK 420
QY 681 SNKPNEOTSXKRDSDTPELKLNTNAPGFTKCSNTSELKEFVNPSLPREEKEKLETYKV 740
Db 421 NNTPEVKISEKHATDAPPEVKVTNIPGLNCCSNKSPPEEVPSPGFORETEENLETQV 480
QY 741 SNAEDPKDLMLSGERVLOQTERSSESSISLVPQTDYCTQESISLLEVSTLQKAKTEPNK 800
Db 481 SNSTKDKNLVLSAEKGLPPERSTESTSVSLVPDQYDQDSISLLETNLRKAKTMSQ 540
QY 801 CVSOCAAFENPKGLIHGCSKDNNDTEGFKYPLGHEVNSHRSSTSEMESELDQYLONT 860
Db 541 CMTQYVAIEHPKELSHSCSKDTRNDTKGFKDALRREVNHIQETNTEMESELDQYLONT 600
QY 861 FKYSKRQSFALFSPGNAAEBCATFSAHSGSLKKQSPKVTFFCEQKEENOGKNESNIKPV 920
Db 601 FOAKSRQSFALFSPNRNPEKEC-----AHSMSLRQSKTKVTLEYEQKEENOGKEESKIKPA 656
QY 921 QTVNITAGFPVVGOKDKPVDNAKCSIKGGSRCFLSSQFRGNETGLITPNKHGGLQNPYRI 980
Db 657 QEVTNTVGFPPVCOEDKPCADANHSITGVSRCLPCLSQLDTNETELSTVKNKHGVLPNPLI 716
QY 981 PPLFPPIKSVKTKCKKNLLEENFEHSMSPEREMGNEN-IPSTVSTISRNNIRENVFKEA 1039
Db 717 PSISPVKSFITTKYKSPSEBEKFKYSASHENAMGNESIIQSIIVYTIQGNIRENAFKEA 776
QY 1040 SSSNINEVSGSTNEVGSINIEIGSSDENIOAELGRNRGPKLAMLRLGVLOPEVYKQSLP 1099
Db 777 SS-----GSINEVSGSGENMQVELGRNRGPKLNSVQRLGHIQIQRKQSLP 822
QY 1100 GSNCKHPEIKK-QEYEEVQVQTNDFSPYILSONLEOPMGSSHASQVCSSETPDDLLDGE 1158
Db 823 VSS-KHEPMKKQENEGVQAVDAFSPCLTSNSLEQPMRSGCVQICSETPDDLLDDKE 881
QY 1159 IKEDTSAENDIKESSAVFSKQVKGELSRSPFTTHLAQCYRRGAKKLESSEENLSS 1218
Db 1159 IKEDTSAENDIKESSAVFSKQVKGELSRSPFTTHLAQCYRRGAKKLESSEENLSS 1218
```



```
Db 882 VKENTSTEGGIKETS AVFSKSTORGESSRSPSLTHISLQAQCHRRARKLESSESISS 941
QY 1219 EDEELPCFQHLLECKVNNIPQSQRHSTVATECLSKNTEENLLSLKNSLDCSNQVILAK 1278
Db 942 EDEELPCFQHLLECKVNTTSQSRGRNAVTEYLSKTKENLVSLKTLTDLDCNDLLILAK 1001
QY 1279 ASQEHHLSEETKSCASLFSQCSSELEDLTANTNTQDPFLI--GSSQMRHQSQSGVGLS 1336
Db 1002 TSOEPLHSEBAKYSGSLFSQCSVLEDLPANTNTQSFLMFGPPSKQTRHQSNEQEVLS 1061
QY 1337 DK 1338
Db 1062 DK 1063

RESULT 13
Q8WMT3 PRELIMINARY; PRT; 947 AA.
AC Q8WMT3;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE BRCA1 (Fragment).
CN BRCA1.
OS Tapirus pinchaque (Mountain tapir).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Tapiridae; Tapirus.
OX NCBI_TaxID=30582;
RN [1]
RP SEQUENCE FROM N.A.
RA Murphy W.J., Eizirik E., O'Brien S.J., Madsen O., Scally M., Douady C.,
RA Teeling E., Ryder O., Stanhope M.J., deJong W.W., Springer M.S.;
RT "Resolution of the early placental mammal radiation using Bayesian
RT Phylogenetics.";
RL Science 0-0-0(2001).
DR EMBL; AY057830; AAL30077.1; -.
FT NON_TER 1
FT NON_TER 947
SQ SEQUENCE 947 AA; 104734 MW; 21D8DE08EBDE0D43 CRC64;

Query Match 36.9%; Score 3557.5; DB 6; Length 947;
Best Local Similarity 75.1%; Pred. No. 1.4e-175;
Matches 713; Conservative 79; Mismatches 152; Indels 5; Gaps 5;

QY 274 CGTNTHASSLQHNSSLLLTKDRMNYEKAFCNKSQKOPGLARSOHNRWAGSKETCNDRT 333
Db 1 CGTNTHASSLQHNSSLLLTKDRMNYEKAFCNKSQKOPGLARSOHNRWAGSKETCNDRT 60
QY 334 PSTKKVVDLADPLCERKEKWNKOKLPCSENPRDTEVPWITLNSSIQKVNWFERSDELL 393
Db 61 PNSEKKVLVNLADPLHGRKELNKKQPCSNSSRSDQDVPWITLNSSIQKVNWFERSDEML 120
QY 394 GSDSDHGESENAKVADVLVDLNEVDYSGSEKIDLLASDPHEALICKSERVHKSVE 453
Db 121 SDDSDRGGPTEHVAGAVEVNEVDYSGSEKIDLMASDPGALICESERVHKSQVPA 180
QY 454 SNIEDKIFGTYRKASPLNSLHWNTENLIICAFVTEPQIIFQERPLTNKLRKRRTPSGLH 513
Db 181 NNIEDKIFGTYRKANPLNSLHSHITELIIEASAIEPQIIFQERPLTNKLRKRRTPSGLH 240
QY 514 PEDFIKKADLA-VQKPEMTNOGTNOTEQNGQVNMNTNSGHENKTKGDSIQNEKNPNPIE 572
Db 241 PEDFIKKVDFAIQVQKPEKVEGTEQIEQNDHVMNTANNHENETKGCVOKEKNANPTE 300
QY 573 SLEKSAFKTKAEPISSSISNMELELNHNSKAPKNLRKRKSTRHIALELVVSRNLS 632
Db 301 SLEKSAFKTEAPEPISSISNLELNHNSKAPKNLRKRKSTRHIALELVVSRNPS 360
QY 633 PPNCTELQIDCCSSSEIEIKKKYKNOMPVRHSRLQLMGEGKEPATGAKKSNKPNEQTSKRH 692
Db 361 PPNHTELOIDCCSSSEIEIKKKNSQGPGRHRRKLQLMEDKAPATGAKKSNKPNEQINKRL 420
QY 693 DSDTFPELKLTNAGSFTKCSNTSELKFYNPSLPREKEEKLETYKVSNNADDPKDLML 752
```

```
Db 421 ASHAFPELKLNTNPGFANCSSNKLQEFVNPSLQREETEQNLGATQVFNSSKDPKDLIL 480
QY 753 SGERVLTQTERSESSISLVPGDYGTQESISLVEYSTLGKAKTEPNKCVSOCAAFENPK 812
Db 481 SSGRGLQAEKRSVESTISLVDPDYGTQDSISLLEADPLGKAKTAPNAGASLCAATENSK 540
QY 813 GLTHGCSKNRNDTEGKYPLGHEVNHRSRETSLEMESELDQAQYLQNTFKVSKRQSFALF 872
Db 541 ELMHGCSKOTRNDTEGLKDPLEXRVNHTQEPSTIEMESELDTOYLHNTFKVSKRQSFALF 600
QY 873 SNPGNAEECATFSAGSLKQSPKVTPECBOKEKNQGNESNIKPQVTNITAGFPVV 932
Db 601 SNSGN-PEECATVCAHSRPLRKESPKVTLECGRKEENQGNKESKIKHVQSVHTPAGEFPVV 659
QY 933 GOKDKPVDNAKCSIKGSRFLSSQPRGNETGLITPNKKHLLQNPYRIPPLPKISFVKT 992
Db 660 CQDKPDGVYKSLKVRSCQSPRGNETELITPNKKHLLQNPYRIPPLPKISFVKT 719
QY 993 KCKKNLLEENFEHSMSPREMGNEN-IPSTVSTISNNIRENVFKEASSSNIINEVGSST 1051
Db 720 TCKKNVSEKLEEHVSYPERAXGNESIIOSTNSTISQNNIRESTFEVGSSTINEVGSST 779
QY 1052 NEVGSSINEIGSSDENIQAEELGRNRPKNAMLRVLQPEVYKQSLPGSNCKHPKIKO 1111
Db 780 NEVGSSINEVSSGENIQAEELGRNRPKNAILRLGLMQPEVYKQSHPISNCKHPKIKRQ 839
QY 1112 -EYEEVVQVTNTDFSPYLISDNLEQPMGSSHASQVCSPTDDLLDGEIKEDETSFAENDI 1170
Db 840 GEMEGVQAVNTDFSPCLISDNLEQPMGSS-CAQVCSPTDDLLDNDKIEENINFAESGI 998
QY 1171 KESSAVFSKSVQKGLSRSPSPETHLAQGYRRGAKKLESSEENLSSE 1219
Db 899 KERSAVFSKSVQKGFGRSPSPAHTCLAQGHORGARKLESSEENMSSE 947

RESULT 14
Q9BDV3 PRELIMINARY; PRT; 947 AA.
AC Q9BDV3;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE BRCA1 (Fragment).
OS Diceros bicornis (Black rhinoceros).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Rhinocerotidae; Diceros.
OX NCBI_TaxID=9805;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21082081; PubMed=11214318;
RA Madsen O., Scally M., Douady C., Kao D., DeBry R., Adkins R.,
RA Amrine H., Stanhope M., de Jong W., Springer M.;
RT "Parallel adaptive radiations in two major clades of placental
RT mammals.";
RL Nature 409:610-614(2001).
DR EMBL; AF284011; AAK15598.1; -.
DR InterPro: IPR002378; Brst_cancer1.
DR PRINTS; PR00493; BRSTCANCER1.
FT NON_TER 1
FT NON_TER 947
SQ SEQUENCE 947 AA; 104753 MW; BC8158CB0F9A404B CRC64;

Query Match 36.8%; Score 3555.5; DB 6; Length 947;
Best Local Similarity 75.2%; Pred. No. 1.4e-175;
Matches 714; Conservative 86; Mismatches 143; Indels 7; Gaps 5;

QY 274 CGTNTHASSLQHNSSLLLTKDRMNYEKAFCNKSQKOPGLARSOHNRWAGSKETCNDRT 333
Db 1 CGTNTHASSLQHNSSLLLTKDRMNYEKAFCNKSQKOPGLARSOHNRWAGSKETCNDRT 60
QY 334 PSTKKVVDLADPLCERKEKWNKOKLPCSENPRDTEVPWITLNSSIQKVNWFERSDELL 393
Db 61 PNSEKKVLVNLADPLHGRKELNKKQPCSNSSRSDQDVPWITLNSSIQKVNWFERSDEML 120
QY 394 GSDSDHGESENAKVADVLVDLNEVDYSGSEKIDLLASDPHEALICKSERVHKSVE 453
Db 121 SDDSDRGGPTEHVAGAVEVNEVDYSGSEKIDLMASDPGALICESERVHKSQVPA 180
QY 454 SNIEDKIFGTYRKASPLNSLHWNTENLIICAFVTEPQIIFQERPLTNKLRKRRTPSGLH 513
Db 181 NNIEDKIFGTYRKANPLNSLHSHITELIIEASAIEPQIIFQERPLTNKLRKRRTPSGLH 240
QY 514 PEDFIKKADLA-VQKPEMTNOGTNOTEQNGQVNMNTNSGHENKTKGDSIQNEKNPNPIE 572
Db 241 PEDFIKKVDFAIQVQKPEKVEGTEQIEQNDHVMNTANNHENETKGCVOKEKNANPTE 300
QY 573 SLEKSAFKTKAEPISSSISNMELELNHNSKAPKNLRKRKSTRHIALELVVSRNLS 632
Db 301 SLEKSAFKTEAPEPISSISNLELNHNSKAPKNLRKRKSTRHIALELVVSRNPS 360
QY 633 PPNCTELQIDCCSSSEIEIKKKYKNOMPVRHSRLQLMGEGKEPATGAKKSNKPNEQTSKRH 692
Db 361 PPNHTELOIDCCSSSEIEIKKKNSQGPGRHRRKLQLMEDKAPATGAKKSNKPNEQINKRL 420
QY 693 DSDTFPELKLTNAGSFTKCSNTSELKFYNPSLPREKEEKLETYKVSNNADDPKDLML 752
```

```
Dh 61 PNSEKLVNADPLVYGRKELNKKQPCSDSPRDSQDIPWITRNSIQKVNWFSSDEIL 120
QY 394 GSDSDHSGESNAKVAADVLDVLNEVDYSGSSEKIDLLASDPHEALICKSRVHSKSV 453
Dh 121 TSDDSHDGGPESNTEVAGAVEQNEVDYSGSSEKIGLWASDPGALICESERVHSKPVE 180
QY 454 SNIEDKIFGKYRKKASLPNLNLSHTVENLIIGAFVTEPQIIQERPLTNLKKRRRPTSG 513
Dh 181 NNTEDKIFGKYRKKASLPNLNLSHTEDLIIGASATESQITQERPLTNLKKRRRPTSG 240
QY 514 PEDFIKADLA-VOKTPEMINOGTNOEONGOVNITNSGHNKTKGDSIQNEKNPNPIE 572
Dh 241 PEDFIKADLVVQKTPKIIIEGDIQIEQNGRVMGIANNHNETKGYVQKEKNANPTE 300
QY 573 SLEKESAFKTKAPPISSISNMELELNHNSKAPKKNRRLRRKSSRHHIALELVYSRNL 632
Dh 301 SLEKESAFKTKAPPISSISNMELELNHNSKAPKKNRRLRRKSSRHHIALELVYSRNL 360
QY 633 PPNCTELQIDSCSSSEIEKKKYNQMPVHRHNSRNLQIMEGKEPATGAKSKNPNQTSKR 692
Dh 361 PPNHTELQVDSCSSSEEMKKSPSPVPVHRHRSKQLQITDEKPAACAKSKNPNQINRKL 420
QY 693 DSDTFPELKLITNAPGFTKCSNTSELKEFVNPSLPREKEEKLTVKVSNNNAEDPKDML 752
Dh 421 ASDAFPELKLITNAPGFTKCSNTSELKEFVNPSLPREKEEKLTVKVSNNNAEDPKD 480
QY 753 SGERVLOTERSVSSISLVPGDTGTOESISLLEVLSTLGRKAKTEPNKCVSOCAAFENPK 812
Dh 481 SGERVLOTERSVSSISLVPGDTGTOESISLLEVLSTLGRKAKTEPNKCVSOCAAFENPK 540
QY 813 GLIHGCSKDNRRNDEGFKYPLGHEVNHRSRETSIEMEESELDQAYLQNTFKVSKRSFALF 872
Dh 541 ELTHDCSKDTRNDEGFKYPLGHEVNHRSRETSIEMEESELDQAYLQNTFKVSKRSFALF 600
QY 873 SNGPNAEECATFSAHSGSLAKQSPKVTFFCEOEKNEOGKNEKNIKPVQTNITAGFPV 932
Dh 601 SNGPNAEECATFSAHSGSLAKQSPKVTFFCEOEKNEOGKNEKNIKPVQTNITAGFPV 660
QY 933 GQKD-KPVDNAKCSIKGSRFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLFIKSPVK 991
Dh 661 CQDKKPGDVKYKCSKTEASRLCQSSQFRGNETGLITPNKHGLLQNPYRIPPLFIKSPVK 720
QY 992 TKCKKNLEENFEHSMSPERMGHENI-IPSTVSTISRNNTRENVFKEASSNINEVGSS 1050
Dh 721 TCKKNVSEKLEHSMSPERMGHENI-IPSTVSTISRNNTRENVFKEASSNINEVGSS 780
QY 1051 TNEVGSSINEGSSDENIQAEGLGRNGPKLNAIIRLGLMQPEVYKQSLPGSNCKHPEIK 1110
Dh 781 TNEVGSSINEGSSDENIQAEGLGRNGPKLNAIIRLGLMQPEVYKQSLPGSNCKHPEIK 840
QY 1111 Q-EYEWVOTVNTDFSPYILISDNLEQPMGSSHASOVCSFETPDDLDDGEIKEDTSFAED 1169
Dh 841 RGENGVQPVNADSPCIPISNLEQPMGSSHASOVCSFETPDDLDDGEIKEDTSFAED 900
QY 1170 IKESSAVFSKQVKGELSPSPFTHLAQYRRGAKKLESSEENLSSE 1219
Dh 901 IKERSAIFSQVKGELSPSPFTHLAQYRRGAKKLESSEENLSSE 947

RESULT 15
Q9BDV4
ID Q9BDV4 PRELIMINARY; PRT: 942 AA.
AC Q9BDV4;
DT 01-JUN-2001 (TreeBLrel. 17, Created)
DT 01-JUN-2001 (TreeBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TreeBLrel. 18, Last annotation update)
DE BRCA1 (Fragment).
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RX MEDLINE=21082081; PubMed=11214318;
RA Madsen O., Scally M., Douady C., Kao D., DeBry R., Adkins R.,
RA Amrine H., Stanhope M., de Jong W., Springer M.;
RT "Parallel adaptive radiations in two major clades of placental
mammals.";
RL Nature 409:610-614(2001).
DR EMBL: AF284010; AAK15597.1; -.
DR InterPro: IPR002378; Brst_CancerI.
DR PRINTS: PR00493; BRSTCANCERI.
FT NON_TER 942 942
FT NON_TER 1 1
SQ SEQUENCE 942 AA; 104317 MW; F5C2EEEIC7313C1A CRC64;

Query Match 36.1%; Score 3482; DB 6; Length 942;
Best Local Similarity 73.6%; Pred. No. 9e-172;
Matches 698; Conservative 90; Mismatches 151; Indels 10; Gaps 4;

QY 274 CGTNTHASSLQHENSSLLLTKDRMNEKAEFCNKSQKPLARSOHNRWAGSKETCNRRT 333
Dh 1 CGTNTHASSLQHENSSLLLTKDRMNEKAEFCNKSQKPLAKSQOSRAESKETCNRQA 60
QY 334 PSTEKKYVDLNDPCLCEKERNKOKLPCSENPRDTEDEVPWITLNSSIQKVNWFSSDELL 393
Dh 61 PNSEKLVNADPLVYGRKELNKKQPCSDSPRDSQDIPWITLNSSIQKVNWFSSDELL 120
QY 394 GSDSDHSGESNAKVAADVLDVLNEVDYSGSSEKIDLLASDPHEALICKSRVHSKSV 453
Dh 121 TSDDSHDGGPESNTEVAGAVEQNEVDYSGSSEKIGLWASDPGALICESERVHSKPVE 180
QY 454 SNIEDKIFGKYRKKASLPNLNLSHTVENLIIGAFVTEPQIIQERPLTNLKKRRRPTSG 513
Dh 181 NNTEDKIFGKYRKKASLPNLNLSHTEDLIIGASATESQITQERPLTNLKKRRRPTSG 240
QY 514 PEDFIKADLA-VOKTPEMINOGTNOEONGOVNITNSGHNKTKGDSIQNEKNPNPIE 572
Dh 241 PEDFIKADLVVQKTPKIIIEGDIQIEQNGRVMGIANNHNETKGYVQKEKNANPTE 300
QY 573 SLEKESAFKTKAPPISSISNMELELNHNSKAPKKNRRLRRKSSRHHIALELVYSRNL 632
Dh 301 SLEKESAFKTKAPPISSISNMELELNHNSKAPKKNRRLRRKSSRHHIALELVYSRNL 360
QY 633 PPNCTELQIDSCSSSEIEKKKYNQMPVHRHNSRNLQIMEGKEPATGAKSKNPNQTSKR 692
Dh 361 PPNHTELQVDSCSSSEEMKKSPSPVPVHRHRSKQLQITDEKPAACAKSKNPNQINRKL 420
QY 693 DSDTFPELKLITNAPGFTKCSNTSELKEFVNPSLPREKEEKLTVKVSNNNAEDPKDML 752
Dh 421 ASDAFPELKLITNAPGFTKCSNTSELKEFVNPSLPREKEEKLTVKVSNNNAEDPKD 480
QY 753 SGERVLOTERSVSSISLVPGDTGTOESISLLEVLSTLGRKAKTEPNKCVSOCAAFENPK 812
Dh 481 SGERVLOTERSVSSISLVPGDTGTOESISLLEVLSTLGRKAKTEPNKCVSOCAAFENPK 540
QY 813 GLIHGCSKDNRRNDEGFKYPLGHEVNHRSRETSIEMEESELDQAYLQNTFKVSKRSFALF 872
Dh 541 ELTHDCSKDTRNDEGFKYPLGHEVNHRSRETSIEMEESELDQAYLQNTFKVSKRSFALF 600
QY 873 SNGPNAEECATFSAHSGSLAKQSPKVTFFCEOEKNEOGKNEKNIKPVQTNITAGFPV 932
Dh 601 SNGPNAEECATFSAHSGSLAKQSPKVTFFCEOEKNEOGKNEKNIKPVQTNITAGFPV 660
QY 933 GQKD-KPVDNAKCSIKGSRFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLFIKSPVK 991
Dh 661 CQDKKPGDVKYKCSKTEASRLCQSSQFRGNETGLITPNKHGLLQNPYRIPPLFIKSPVK 720
QY 992 TKCKKNLEENFEHSMSPERMGHENI-IPSTVSTISRNNTRENVFKEASSNINEVGSS 1050
Dh 721 TCKKNVSEKLEHSMSPERMGHENI-IPSTVSTISRNNTRENVFKEASSNINEVGSS 774
QY 1051 TNEVGSSINEGSSDENIQAEGLGRNGPKLNAIIRLGLMQPEVYKQSLPGSNCKHPEIK 1110
Dh 775 -NEVGSSINEGSSGEHTQAEGLGRNGPKLNAIIRLGLMQPEVYKQSLPGSNCKHPEIK 833
```

QY 1111 QEYEEVQTVNTDFSPYLIISDNLEQPMGSSHASQVCSETPDDLLDDGGEIKEDTSFAENDI 1170
| : |||| || |||| |||| |||| |||| |||| |||| |||| : |||| :
Db 834 QGEKEVYQAYNADEFPCLIISDNLEQPMGSSCASQVCSETPDDLLINDDEIKENISFAESGV 893
QY 1171 KESSAVFSKSVQKGELSRSPFTHTHLAQGYRRGAKKLESSSENLSSE 1219
|| |||| |||| : |||| | |||| : |||| : |||| |||| ||||
Db 894 KERSAVFSKSVQKGKFRFRSPIGRTCLAQGHORRARKLESSENTSSE 942

Search completed: January 22, 2003, 16:48:27
Job time : 83 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 22, 2003, 16:46:08 ; Search time 18 Seconds
(without alignments)
3045.272 Million cell updates/sec

Title: US-09-734-672-4

Perfect score: 9649

Sequence: 1 MDLSALRVEEVQNVINAMOK.....LYQCQLDYLPQIPHSHV 1863

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 segs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2.6/ptodata/1/1aa/5A_COMB.pep.*

2: /cgn2.6/ptodata/1/1aa/5B_COMB.pep.*

3: /cgn2.6/ptodata/1/1aa/6A_COMB.pep.*

4: /cgn2.6/ptodata/1/1aa/6B_COMB.pep.*

5: /cgn2.6/ptodata/1/1aa/PCTUS_COMB.pep.*

6: /cgn2.6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9649	100.0	1863	1 US-08-798-691-4	Sequence 4, Appl
2	9649	100.0	1863	3 US-08-825-487A-4	Sequence 4, Appl
3	9649	100.0	1863	4 US-09-074-476-6	Sequence 6, Appl
4	9642	99.9	1863	1 US-08-480-784-2	Sequence 2, Appl
5	9642	99.9	1863	1 US-08-483-553-2	Sequence 2, Appl
6	9642	99.9	1863	1 US-08-487-002-2	Sequence 2, Appl
7	9642	99.9	1863	1 US-08-483-554B-2	Sequence 2, Appl
8	9642	99.9	1863	1 US-08-488-011B-2	Sequence 2, Appl
9	9642	99.9	1863	4 US-08-850-727-2	Sequence 2, Appl
10	9642	99.9	1863	5 PCT-US95-10202-2	Sequence 2, Appl
11	9642	99.9	1863	5 PCT-US95-10203-2	Sequence 2, Appl
12	9642	99.9	1863	5 PCT-US95-10220-2	Sequence 2, Appl
13	9635	99.9	1863	1 US-08-598-591-2	Sequence 2, Appl
14	9635	99.9	1863	1 US-08-798-691-2	Sequence 2, Appl
15	9635	99.9	1863	1 US-08-798-691-6	Sequence 6, Appl
16	9635	99.9	1863	3 US-08-825-487A-2	Sequence 2, Appl
17	9635	99.9	1863	3 US-08-825-487A-6	Sequence 6, Appl
18	9635	99.9	1863	4 US-09-074-476-2	Sequence 2, Appl
19	9635	99.9	1863	4 US-09-074-476-4	Sequence 4, Appl
20	9630	99.8	1863	1 US-08-425-061-16	Sequence 16, Appl
21	9630	99.8	1863	2 US-08-825-886-16	Sequence 16, Appl
22	9581	99.3	1863	2 US-08-603-753D-2	Sequence 2, Appl
23	9581	99.3	1863	4 US-09-099-753-2	Sequence 2, Appl
24	9581	99.3	1863	4 US-08-986-106-2	Sequence 2, Appl
25	9581	99.3	1863	4 US-09-007-678B-49	Sequence 49, Appl
26	9577	99.3	1852	1 US-08-425-061-24	Sequence 24, Appl
27	9577	99.3	1852	2 US-08-825-886-24	Sequence 24, Appl

28	6997	72.5	1363	1 US-08-425-061-23	Sequence 23, Appl
29	6997	72.5	1363	2 US-08-825-886-23	Sequence 23, Appl
30	6222	64.5	1202	1 US-08-425-061-22	Sequence 22, Appl
31	6222	64.5	1202	2 US-08-825-886-22	Sequence 22, Appl
32	4730	49.0	914	1 US-08-425-061-21	Sequence 21, Appl
33	4730	49.0	914	2 US-08-825-886-21	Sequence 21, Appl
34	4621	47.9	900	1 US-08-425-061-20	Sequence 20, Appl
35	4621	47.9	900	2 US-08-825-886-20	Sequence 20, Appl
36	3958	41.0	765	1 US-08-425-061-19	Sequence 19, Appl
37	3958	41.0	765	2 US-08-825-886-19	Sequence 19, Appl
38	1596	16.5	312	1 US-08-425-061-18	Sequence 18, Appl
39	1596	16.5	312	2 US-08-825-886-18	Sequence 18, Appl
40	1563	16.2	309	2 US-08-785-464-5	Sequence 5, Appl
41	1552	16.1	301	2 US-08-785-464-1	Sequence 1, Appl
42	1020	10.6	196	2 US-08-785-464-3	Sequence 3, Appl
43	558	5.8	106	2 US-08-785-464-2	Sequence 2, Appl
44	520	5.4	100	4 US-09-230-196-3	Sequence 3, Appl
45	492	5.1	92	2 US-08-785-464-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-798-691-4
; Sequence 4, Application US/08798691
; Patent No. 5750400
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; APPLICANT: Allen, Antonette C.
; APPLICANT: Alvares, Christopher P.
; APPLICANT: Critz, Brenda S.
; APPLICANT: Olson, Sheri J.
; APPLICANT: Schelter, Denise B.
; APPLICANT: Zeng, Bin
; TITLE OF INVENTION: Coding Sequences of the Human
; TITLE OF INVENTION: BRCA1 Gene
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ONCORMED
; STREET: 200 Perry Parkway
; CITY: Gaithersburg
; STATE: MD
; COUNTRY: USA
; ZIP: 20877
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/798,691
; FILING DATE: 12-Feb-97
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Thomas Gallegos
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: PA-0054CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-527-2051
; TELEFAX: 301-208-6997
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCA1
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17

; MAP POSITION: 17q21									
US-08-798-691-4									
Query Match		100.08; Score 9649; DB 1; Length 1863;							
Best Local Similarity		100.08; Pred. No. 0;							
Matches 1863; Conservative		0; Mismatches 0; Indels 0; Gaps 0;							
QY	1	MDLSALRVEEVQNVINAMQKILECPICLLEIKPEVSTKCDHIFCKFCMLKLLNQKGPQ	60						
DB	1	MDLSALRVEEVQNVINAMQKILECPICLLEIKPEVSTKCDHIFCKFCMLKLLNQKGPQ	60						
QY	61	CPCKNDITKRSLOSTRFPSQVLEBLLKICAFQDQDTGLEVANSYNFAKKENNSPEHLKD	120						
DB	61	CPCKNDITKRSLOSTRFPSQVLEBLLKICAFQDQDTGLEVANSYNFAKKENNSPEHLKD	120						
QY	121	EYSIIQSMGYRRNRKRLLOSEPENSLOFTSLVSQLSNLGTVRTLRTKQRTQPKTSYVI	180						
DB	121	EYSIIQSMGYRRNRKRLLOSEPENSLOFTSLVSQLSNLGTVRTLRTKQRTQPKTSYVI	180						
QY	181	ELGSSDSEDVANKATYCSVGDOELLQITPQGTTRDEISLDSAKAACSEFSDVTNTEHHQ	240						
DB	181	ELGSSDSEDVANKATYCSVGDOELLQITPQGTTRDEISLDSAKAACSEFSDVTNTEHHQ	240						
QY	241	PSNNDLNTTEKRAAERHPEKIQGSSVSNLHVPCGCTNTHASSLQHSNLSLLTKDRMVE	300						
DB	241	PSNNDLNTTEKRAAERHPEKIQGSSVSNLHVPCGCTNTHASSLQHSNLSLLTKDRMVE	300						
QY	301	KAFCNKSQKQGLARQSHNRWAGSKETCNDRTPTSTKVKDLNADPLCERKEWNKQLPC	360						
DB	301	KAFCNKSQKQGLARQSHNRWAGSKETCNDRTPTSTKVKDLNADPLCERKEWNKQLPC	360						
QY	361	SENPRDTEVPWITLNSSTQKWNFWSRDELGGSDSHDGESESNKAVADVLDVLEVD	420						
DB	361	SENPRDTEVPWITLNSSTQKWNFWSRDELGGSDSHDGESESNKAVADVLDVLEVD	420						
QY	421	EYSGSSEKIDLLASDPHEALICKSRVHSKSVESNIEDKIFGKYRKASLPNLSHVTE	480						
DB	421	EYSGSSEKIDLLASDPHEALICKSRVHSKSVESNIEDKIFGKYRKASLPNLSHVTE	480						
QY	481	LIIGAFVTEPQIIQERPLTNLKRKRRTSGLHPDEFKTKADLAVQKTPMINOGTNOTE	540						
DB	481	LIIGAFVTEPQIIQERPLTNLKRKRRTSGLHPDEFKTKADLAVQKTPMINOGTNOTE	540						
QY	541	QNGQVNIITNSGHENKTKGDSIQNEKNPNIESLEKESAFKTKAEPISSSISNMELELNI	600						
DB	541	QNGQVNIITNSGHENKTKGDSIQNEKNPNIESLEKESAFKTKAEPISSSISNMELELNI	600						
QY	601	HNSKAPKNRLRRKSTRIHALELVVSRNLSPNCTELQIDSCSSSEIEIKKKYNOMPV	660						
DB	601	HNSKAPKNRLRRKSTRIHALELVVSRNLSPNCTELQIDSCSSSEIEIKKKYNOMPV	660						
QY	661	RHSRLQIMEGKEPATGAKSKNKPEQTSKRHSDTTPPELKLITNAPGSFTKCSNTSELKE	720						
DB	661	RHSRLQIMEGKEPATGAKSKNKPEQTSKRHSDTTPPELKLITNAPGSFTKCSNTSELKE	720						
QY	721	FYNPSLPREEKEKETVKVSNNAEDPKDLMSGERVLQTERSVESSISLVPCTDYGTO	780						
DB	721	FYNPSLPREEKEKETVKVSNNAEDPKDLMSGERVLQTERSVESSISLVPCTDYGTO	780						
QY	781	ESISLLEVSTLGAKTEPNKCVSQAAPENPKGLIHGCSKDNRNDEGFKYPLGHEYNHS	840						
DB	781	ESISLLEVSTLGAKTEPNKCVSQAAPENPKGLIHGCSKDNRNDEGFKYPLGHEYNHS	840						
QY	841	RETSIEMESELDAQYLQNTFKVSKRQSFALFSPNGNAEECATFSAHSGSLKKQSPKVT	900						
DB	841	RETSIEMESELDAQYLQNTFKVSKRQSFALFSPNGNAEECATFSAHSGSLKKQSPKVT	900						
QY	901	FECEQKEENQGNESNIKPVQTVNITAGFPVVGOKDPVDNAKCSIKGGSFCLSSQFRG	960						
DB	901	FECEQKEENQGNESNIKPVQTVNITAGFPVVGOKDPVDNAKCSIKGGSFCLSSQFRG	960						
QY	961	NETGLITPNKHGLQNPYRIPPLFPKISFVKTCKKNLLEENFEHSMSPEREMGNENIP	1020						
DB	961	NETGLITPNKHGLQNPYRIPPLFPKISFVKTCKKNLLEENFEHSMSPEREMGNENIP	1020						

RESULT 2

US-08-825-487A-4
; Sequence 4, Application US/08825487A
; Patent No. 6048689
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; APPLICANT: White, Marga B.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING VARIATIONS IN POLYNUCLEOTIDE SEQUE
; NUMBER OF SEQUENCES: 110
; CORRESPONDENCE ADDRESS:

DB	961	NETGLITPNKHGLQNPYRIPPLFPKISFVKTCKKNLLEENFEHSMSPEREMGNENIP	1020
QY	1021	STVSTISRNNIRENVFKKASSNNINEVCSSSTNEVGSSINEIGSSDENIQAEIGRNRGPKL	1080
DB	1021	STVSTISRNNIRENVFKKASSNNINEVCSSSTNEVGSSINEIGSSDENIQAEIGRNRGPKL	1080
QY	1081	NAMLRGLVQPEVYKQSLPGSNCKHPEIKKQIEYEVVQTVNTDFSPYILISDNLEQPMGSS	1140
DB	1081	NAMLRGLVQPEVYKQSLPGSNCKHPEIKKQIEYEVVQTVNTDFSPYILISDNLEQPMGSS	1140
QY	1141	HASQVCSSTPDLDLDDGEIKEDTSFAENDIKESSAVFSKSVOKGELSRSRPSFTTHLQ	1200
DB	1141	HASQVCSSTPDLDLDDGEIKEDTSFAENDIKESSAVFSKSVOKGELSRSRPSFTTHLQ	1200
QY	1201	GYRGAKKLESSEENLSEDEELPCFOHLLFGKVNINIPQSQTRHSTVATECLSKNTEENL	1260
DB	1201	GYRGAKKLESSEENLSEDEELPCFOHLLFGKVNINIPQSQTRHSTVATECLSKNTEENL	1260
QY	1261	LSLKNLSLNDSCNQVILAKASQEHHLSEETKCSASLFSQCSELEDLTANTNTQDPFLIGS	1320
DB	1261	LSLKNLSLNDSCNQVILAKASQEHHLSEETKCSASLFSQCSELEDLTANTNTQDPFLIGS	1320
QY	1321	SKMRHOGSEOGVGLSKDELVSDEERTGLEENNQEBSQMSDNLGEAASGESETSVSE	1380
DB	1321	SKMRHOGSEOGVGLSKDELVSDEERTGLEENNQEBSQMSDNLGEAASGESETSVSE	1380
QY	1381	DCSGLSSQSDLLTTQORDTMOHNLKLOQEMAELEAVLEQHGQSQPSNSYPSIISDSSALE	1440
DB	1381	DCSGLSSQSDLLTTQORDTMOHNLKLOQEMAELEAVLEQHGQSQPSNSYPSIISDSSALE	1440
QY	1441	DLRNPESQSTSEKAVLTQKSEYPTISQNEGLSADKPEVSADSTSKNKEPGEVERSSPSK	1500
DB	1441	DLRNPESQSTSEKAVLTQKSEYPTISQNEGLSADKPEVSADSTSKNKEPGEVERSSPSK	1500
QY	1501	CPSLDDRWYHSCSGSLQNRNYPQOEELIKVVDVEEQLEESGPHDLTETSYLPRQDLEG	1560
DB	1501	CPSLDDRWYHSCSGSLQNRNYPQOEELIKVVDVEEQLEESGPHDLTETSYLPRQDLEG	1560
QY	1561	TPYLESIGSLFSDDPESDRAPEARSARVGNIPSSTSALKVPOLKVAESQAQSAHAHTT	1620
DB	1561	TPYLESIGSLFSDDPESDRAPEARSARVGNIPSSTSALKVPOLKVAESQAQSAHAHTT	1620
QY	1621	DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEEFMLVYKFARKHHTLTNLI	1680
DB	1621	DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEEFMLVYKFARKHHTLTNLI	1680
QY	1681	TEETHVVMKTDABFVCERTLKYFLGIAGGKVVVSYFWVTQSIKERKMLNEHDFEVRGDV	1740
DB	1681	TEETHVVMKTDABFVCERTLKYFLGIAGGKVVVSYFWVTQSIKERKMLNEHDFEVRGDV	1740
QY	1741	VNGRNHOGPKRARESODRKIFRGLEICCYGPTNNPTDQLEMMVOLCGASVVKELSSFTL	1800
DB	1741	VNGRNHOGPKRARESODRKIFRGLEICCYGPTNNPTDQLEMMVOLCGASVVKELSSFTL	1800
QY	1801	GTGVHPVIVVQPDQAWTENDNGFHAIGQCEAPVTRREWVLDVSVALYQCELODTYLIQPIPH	1860
DB	1801	GTGVHPVIVVQPDQAWTENDNGFHAIGQCEAPVTRREWVLDVSVALYQCELODTYLIQPIPH	1860
QY	1861	SHY 1863	
DB	1861	SHY 1863	

ADDRESSEE: Howrey & Simon
STREET: 1299 Pennsylvania Avenue., N.W.
CITY: Washington,
STATE: DC
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/825,487A
FILING DATE: 28-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US98/060002
FILING DATE: 26-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Albert P. Halluin
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: 05371.0012.999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-463-8100
TELEFAX: 650-463-8400
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
STRAIN: BRCAL
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 17
MAP POSITION: 17q21
US-08-825-487A-4

Query Match 100.0%; Score 9649; DB 3; Length 1863;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDLSALRVEEVQVNVINAMQKILECPICLELIKPEVSTKCDHIFCKFCMLKLLNQKGPQS 60
DB 1 MDLSALRVEEVQVNVINAMQKILECPICLELIKPEVSTKCDHIFCKFCMLKLLNQKGPQS 60

QY 61 CPLCKNDITKRSLOESTRFSQVVEELLKIICAFQDGTGLEAYANSYNFAKKENNSPEHLKD 120
DB 61 CPLCKNDITKRSLOESTRFSQVVEELLKIICAFQDGTGLEAYANSYNFAKKENNSPEHLKD 120

QY 121 EVSIIOSMGYRNRAKRLLOSEPENPSLOETSLSVQLNSLGTVRLTKRIQPKTSVYI 180
DB 121 EVSIIOSMGYRNRAKRLLOSEPENPSLOETSLSVQLNSLGTVRLTKRIQPKTSVYI 180

QY 181 ELGSDSSEDTVKNATYCSVGDQELLQITPQGTREISLDSAKKAACEFSETDVTNTEHQ 240
DB 181 ELGSDSSEDTVKNATYCSVGDQELLQITPQGTREISLDSAKKAACEFSETDVTNTEHQ 240

QY 241 PSNNDLNTTEKRAAERHPEKYQSSVSNLHVPCGNTTHASSLQHENSILLTKDRMNYE 300
DB 241 PSNNDLNTTEKRAAERHPEKYQSSVSNLHVPCGNTTHASSLQHENSILLTKDRMNYE 300

QY 301 KAEFCNKSQPLARQSHRNWAGSKETCNDRRTPSTPEKKVVDLNAADPLCERKENNOKLPC 360
DB 301 KAEFCNKSQPLARQSHRNWAGSKETCNDRRTPSTPEKKVVDLNAADPLCERKENNOKLPC 360

QY 361 SENPROTEVPWITLNSSTOKYNEWFSRSDLLGSDSDHGESESNKAVADVLVDLYNEVD 420
DB 361 SENPROTEVPWITLNSSTOKYNEWFSRSDLLGSDSDHGESESNKAVADVLVDLYNEVD 420

QY 421 EYSGSSEKIDLLASDPHEALICKSERVSHKSVESNIEDKIFGKTYRKKASLPNLSHV TEN 480
DB 421 EYSGSSEKIDLLASDPHEALICKSERVSHKSVESNIEDKIFGKTYRKKASLPNLSHV TEN 480

QY 481 LIIGAFTVTEPQIIQERPLTNKLRKRPTSGLHPEDFIKADLAVOKTPEMNOGNTQTE 540
DB 481 LIIGAFTVTEPQIIQERPLTNKLRKRPTSGLHPEDFIKADLAVOKTPEMNOGNTQTE 540

QY 541 QNGOVNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINMELELNI 600
DB 541 QNGOVNITNSGHENKTGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINMELELNI 600

QY 601 HNSKAPKNRLRRKSSSTRHIALELVVSRNLSPPNCTELQIDSCSSESEIKKKYQMPV 660
DB 601 HNSKAPKNRLRRKSSSTRHIALELVVSRNLSPPNCTELQIDSCSSESEIKKKYQMPV 660

QY 661 RHSRNLQMEGKEPATGAKKSNKPNQOTSKRHDSOTFPELKLTNAPGSKTCSNTSELKE 720
DB 661 RHSRNLQMEGKEPATGAKKSNKPNQOTSKRHDSOTFPELKLTNAPGSKTCSNTSELKE 720

QY 721 FVNPSLPREEKEKLETVKVSNNADPKDMLSGERVLTQTERSVESSISLVPGTDYGTQ 780
DB 721 FVNPSLPREEKEKLETVKVSNNADPKDMLSGERVLTQTERSVESSISLVPGTDYGTQ 780

QY 781 ESISLLEVSTLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDRNDTEGFKYPLGHEVNH 840
DB 781 ESISLLEVSTLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDRNDTEGFKYPLGHEVNH 840

QY 841 RETSIEMESELDAQYLQNTFKVSRQSFALPSNPGNAEEECATFSAHSGSLKKQSPKVT 900
DB 841 RETSIEMESELDAQYLQNTFKVSRQSFALPSNPGNAEEECATFSAHSGSLKKQSPKVT 900

QY 901 FECEQKEENOGKNESNIKPVQTVNITAGFPVVGOKDPVDNAKCSIKGGSRCFLSSQPRG 960
DB 901 FECEQKEENOGKNESNIKPVQTVNITAGFPVVGOKDPVDNAKCSIKGGSRCFLSSQPRG 960

QY 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVTKCKKNLLEENFEEHSMSPEREMGNENIP 1020
DB 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVTKCKKNLLEENFEEHSMSPEREMGNENIP 1020

QY 1021 STVSTISRNRIENVEFKEASSNINEVSGSTNEVGSSINEIGSSDENIOAELGRNRPKL 1080
DB 1021 STVSTISRNRIENVEFKEASSNINEVSGSTNEVGSSINEIGSSDENIOAELGRNRPKL 1080

QY 1081 NAMRLGVLOQPEVYKOSLPGSNCKHPEIKKOEVEVQVNTDFSPYLISDNLEQPMGSS 1140
DB 1081 NAMRLGVLOQPEVYKOSLPGSNCKHPEIKKOEVEVQVNTDFSPYLISDNLEQPMGSS 1140

QY 1141 HASQVCSETPDDLLDDGEIKEDTSPAENDIKESSAVFSKSVQKGLSRSPSPFTHLQAQ 1200
DB 1141 HASQVCSETPDDLLDDGEIKEDTSPAENDIKESSAVFSKSVQKGLSRSPSPFTHLQAQ 1200

QY 1201 GYRRGAKKLESSEENLSEDEELPCFQHLFGKVNNIPSOQSTRHSTVATECLSKTEENL 1260
DB 1201 GYRRGAKKLESSEENLSEDEELPCFQHLFGKVNNIPSOQSTRHSTVATECLSKTEENL 1260

QY 1261 LSLKNSLDCSNQVILAKASQEHHLSEETKCSASLFSOCSSELEDLTANTNTQDPFLIGS 1320
DB 1261 LSLKNSLDCSNQVILAKASQEHHLSEETKCSASLFSOCSSELEDLTANTNTQDPFLIGS 1320

QY 1321 SKOMRHQSESQGVGLSDKELVSDDEERGTLREENQOEQSDSNLGEAASGCESETSVSE 1380
DB 1321 SKOMRHQSESQGVGLSDKELVSDDEERGTLREENQOEQSDSNLGEAASGCESETSVSE 1380

QY 1381 DCSGLSSQSDIITTOQRTMQHNLKLOEMAELAVLEBQHSQPSNSYPSIISDSSALE 1440
DB 1381 DCSGLSSQSDIITTOQRTMQHNLKLOEMAELAVLEBQHSQPSNSYPSIISDSSALE 1440

QY 1441 DLRNPEQSTSEKAVLTSSOKSEVPISQNPESADKFEVSADSSSTSKNKEPVERSSPSK 1500
DB 1441 DLRNPEQSTSEKAVLTSSOKSEVPISQNPESADKFEVSADSSSTSKNKEPVERSSPSK 1500

QY 1501 CPSLDDRWYMHSCSGSLQNRNYPQBELIKVVDVBEQOESESGPHDLTSTSYLPRQDLEG 1560

```
Db 1501 CFSLDRLWYMHSCSGSLQNRNYPQSEELIKVVDVEEQLEESGPHDLTETSYLRQDLEG 1560
QY 1561 TPYLSGSLFSDPSPEDRAPESARVGNIPSTTSALKVPQLKVAESAQSPAAHHT 1620
Db 1561 TPYLSGSLFSDPSPEDRAPESARVGNIPSTTSALKVPQLKVAESAQSPAAHHT 1620
QY 1621 DRAGYNAMESYSREKPELTASTERVNKRMSMVVSGLTPEEFMLVYKFAKHHITLNL 1680
Db 1621 DRAGYNAMESYSREKPELTASTERVNKRMSMVVSGLTPEEFMLVYKFAKHHITLNL 1680
QY 1681 TEETHVVMKTDAEFVCERTLKYFLGIAGGKVVSYFWVTQSIKERKMLNHDHFEVRGDV 1740
Db 1681 TEETHVVMKTDAEFVCERTLKYFLGIAGGKVVSYFWVTQSIKERKMLNHDHFEVRGDV 1740
QY 1741 VNGRNHQGPKRARESQRKIFRGLIEICCYGPTNNPTDQLEWVQVLCGASVVKELSF 1800
Db 1741 VNGRNHQGPKRARESQRKIFRGLIEICCYGPTNNPTDQLEWVQVLCGASVVKELSF 1800
QY 1801 GTGVHPVVVQPDAMTNGFHAIGOMCEAPVVTREWVLDVALYQCOELDTYLIPQIPH 1860
Db 1801 GTGVHPVVVQPDAMTNGFHAIGOMCEAPVVTREWVLDVALYQCOELDTYLIPQIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863

RESULT 3
US-09-074-476-6
; Sequence 6, Application US/09074476
; Patent No. 6130322
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; APPLICANT: Allen, Antonette C.
; APPLICANT: Alvares, Christopher P.
; APPLICANT: Critz, Brenda S.
; APPLICANT: Olson, Sheri J.
; APPLICANT: Thurber, Denise
; APPLICANT: Zeng, Bin
; TITLE OF INVENTION: Coding Sequences of the Human
; TITLE OF INVENTION: BRCAL Gene
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue N. W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,476
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/074,453
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Albert P. Halluin
; REGISTRATION NUMBER: 25,227
; REFERENCE/DOCKET NUMBER: 5371.34.US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8109
; TELEFAX: 650-463-8400
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
```

```
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCAL (Om13)
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17
; MAP POSITION: 17q21
US-09-074-476-6

Query Match 100.0%; Score 9649; DB 4; Length 1863;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1863; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDLSARVEEVQNVINAMQKILECPICILELIKPEVSTKCDHIFCKFCMLKLLNKKGPSQ 60
Db 1 MDLSARVEEVQNVINAMQKILECPICILELIKPEVSTKCDHIFCKFCMLKLLNKKGPSQ 60
QY 61 CPLCKNDITKRSLOESTRFVSOLVVEELLKIIICAFQOLDTCLEAYANSYNFAKKENNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRFVSOLVVEELLKIIICAFQOLDTCLEAYANSYNFAKKENNSPEHLKD 120
QY 121 EVSIIQSMGYNRNRAKRLLOSEPENPSLQETSLSVQLSNLGTVRTLRTKORIPOQKTSYVI 180
Db 121 EVSIIQSMGYNRNRAKRLLOSEPENPSLQETSLSVQLSNLGTVRTLRTKORIPOQKTSYVI 180
QY 181 ELGSDSSEDVTNKATYCSVGQDELLQITPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDVTNKATYCSVGQDELLQITPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNNDLNTTEKRAAERPEKYQGVSSVNLHVPCGTNTHASSLOHENSLLLTDKRMNVE 300
Db 241 PSNNDLNTTEKRAAERPEKYQGVSSVNLHVPCGTNTHASSLOHENSLLLTDKRMNVE 300
QY 301 KAEECNKSKOPGLARSQHRNWRAGSKETCNDRTPTSTKVKVDLADPLCERKEWNKQKLP 360
Db 301 KAEECNKSKOPGLARSQHRNWRAGSKETCNDRTPTSTKVKVDLADPLCERKEWNKQKLP 360
QY 361 SENPRDTEVPWITLNSIOKVNEWFSRDELLGSDSDHGESESNKAVADVLOVLNEVD 420
Db 361 SENPRDTEVPWITLNSIOKVNEWFSRDELLGSDSDHGESESNKAVADVLOVLNEVD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKASLPNLSHVTE 480
Db 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKASLPNLSHVTE 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPMINOQTNQTE 540
Db 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPMINOQTNQTE 540
QY 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPITSSISNNELELNI 600
Db 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPITSSISNNELELNI 600
QY 601 HNSKAPKNRLRRKSSRTHALFLVVRNLSPPNCTLOIDSCSSSDEIKKKYNQMPV 660
Db 601 HNSKAPKNRLRRKSSRTHALFLVVRNLSPPNCTLOIDSCSSSDEIKKKYNQMPV 660
QY 661 RHRNRLQMEGKEPATGAKKSNKPNEQTSKRHDSDTPELKLITNAPGSFTKCSNTSELKE 720
Db 661 RHRNRLQMEGKEPATGAKKSNKPNEQTSKRHDSDTPELKLITNAPGSFTKCSNTSELKE 720
QY 721 FVNPSPREEKEKLETIVKVSNNADPKDMLSGERVLTQTSERSVSSSISLVPQTDYGTQ 780
Db 721 FVNPSPREEKEKLETIVKVSNNADPKDMLSGERVLTQTSERSVSSSISLVPQTDYGTQ 780
QY 781 ESISLLEVSITLGRKAKTEPNKCVSOCAAFENPKGLIHGCSKDNNDTGEFKYPLGHEVNH 840
Db 781 ESISLLEVSITLGRKAKTEPNKCVSOCAAFENPKGLIHGCSKDNNDTGEFKYPLGHEVNH 840
QY 841 RETSIEMESELDAQYLQNTFKVSKRQSFAFNSPNGNAEECAFSAHSGSLKKQSPKVT 900
Db 841 RETSIEMESELDAQYLQNTFKVSKRQSFAFNSPNGNAEECAFSAHSGSLKKQSPKVT 900
```


Db 841 RETSIEMEESELDQYQLQNTFKYKRSQSFALFSPNGNAEECATFSAHSGSLKKQSPKVT 900
QY 901 FECEQKEENQKGNESNKPQVTNITAGFPVVGQDKPVDNAKCSIKGGSRFCLSSQFRG 960
Db 901 FECEQKEENQKGNESNKPQVTNITAGFPVVGQDKPVDNAKCSIKGGSRFCLSSQFRG 960
QY 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVKTKCKNLLBENFEHSHMSPEREMGNINIP 1020
Db 961 NETGLITPNKHGLLQNPYRIPPLPIKSFVKTKCKNLLBENFEHSHMSPEREMGNINIP 1020
QY 1021 STVSTISRNIRENVKREASSNINEVSGSTNEVSGSINEIGSSDENIOAELGRNRPKL 1080
Db 1021 STVSTISRNIRENVKREASSNINEVSGSTNEVSGSINEIGSSDENIOAELGRNRPKL 1080
QY 1081 NAMLRGLVQPEYKYQLPGSNCKHPRIKKQYEEVVQTVNTDFSPYLIISDNLEQPMGSS 1140
Db 1081 NAMLRGLVQPEYKYQLPGSNCKHPRIKKQYEEVVQTVNTDFSPYLIISDNLEQPMGSS 1140
QY 1141 HASQVCSETPDDLLDDGEIKEDTSFAENDIKESSAVFSKVQKGEUSRSPSPFTHHLAQ 1200
Db 1141 HASQVCSETPDDLLDDGEIKEDTSFAENDIKESSAVFSKVQKGEUSRSPSPFTHHLAQ 1200
QY 1201 GYRRGAKKLESSRENLSSEDEELPCFOHLLFGKVNNTPSQSTRHSTVATECLSKNTEENL 1260
Db 1201 GYRRGAKKLESSRENLSSEDEELPCFOHLLFGKVNNTPSQSTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDCSNQVILAKASQEHLSBETKCSASLFSQCSSELDLTANTNTODPFLIGS 1320
Db 1261 LSLKNSLNDCSNQVILAKASQEHLSBETKCSASLFSQCSSELDLTANTNTODPFLIGS 1320
QY 1321 SKMRHQSOGVGLSKDELVSDDERGTGLENNQEQSMDSNGLGEAASGCESETSVSE 1380
Db 1321 SKMRHQSOGVGLSKDELVSDDERGTGLENNQEQSMDSNGLGEAASGCESETSVSE 1380
QY 1381 DCSGLSSOSDILITQOQDTHQHNLIKLOQEMAELEAVLEOHGQSPNSYPSIISDSALE 1440
Db 1381 DCSGLSSOSDILITQOQDTHQHNLIKLOQEMAELEAVLEOHGQSPNSYPSIISDSALE 1440
QY 1441 DLNRPEQSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSTSKNKEPVGVERSPSK 1500
Db 1441 DLNRPEQSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSTSKNKEPVGVERSPSK 1500
QY 1501 CPSLDDRWYHSCGSLQNRNYPQSEBELIKVVDVEEQLEESGPHDLTETSYLPRODLEG 1560
Db 1501 CPSLDDRWYHSCGSLQNRNYPQSEBELIKVVDVEEQLEESGPHDLTETSYLPRODLEG 1560
QY 1561 TPYLESGISLFSDDPESDPSDRAPESARVGNIPSPSTSAKLVQPKVAESAQSPAHAHT 1620
Db 1561 TPYLESGISLFSDDPESDPSDRAPESARVGNIPSPSTSAKLVQPKVAESAQSPAHAHT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERNKRMVMVVSGLTPEEFMLVYKFARKHHTLTNLI 1680
Db 1621 DTAGYNAMESVSREKPELTASTERNKRMVMVVSGLTPEEFMLVYKFARKHHTLTNLI 1680
QY 1681 TEETHVVMKTDAEFCERTLKVFLGIAGGKVVVYFWVTQSIKERKMLNEHDFEVRGDV 1740
Db 1681 TEETHVVMKTDAEFCERTLKVFLGIAGGKVVVYFWVTQSIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHOGPKRARESQRDKLFRGLETCCTCGPFTNMPTDQLEWVQCGASVVKELSFLL 1800
Db 1741 VNGRNHOGPKRARESQRDKLFRGLETCCTCGPFTNMPTDQLEWVQCGASVVKELSFLL 1800
QY 1801 GTGVHPPIVVVQPDADWEDNGFHAIGQCEAPVVTREWLDSVALYOCQELDTYLIPQIPH 1860
Db 1801 GTGVHPPIVVVQPDADWEDNGFHAIGQCEAPVVTREWLDSVALYOCQELDTYLIPQIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863

RESULT 4

US-08-480-784-2

; Sequence 2, Application US/08480784
; Patent No. 5693473
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,784
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1963 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-784-2

Query Match 99.9%; Score 9642; DB 1; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDLSALRVEEVQVNVINAMQKILECPICLELIKPEVSTKCDHIFCKFCMLKLLNQKKGPSQ 60
Db 1 MDLSALRVEEVQVNVINAMQKILECPICLELIKPEVSTKCDHIFCKFCMLKLLNQKKGPSQ 60

QY 61 CPLCKNDITKRSLOESTRFSQVLSQVEELLKIICAFQDGTGLGLEYANSYFACKENNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRFSQVLSQVLEELLKIICAFQDGTGLGLEYANSYFACKENNSPEHLKD 120

QY 121 EVSIIOSMGYNRAKRLIOSEPENPSLOETSLVSQLSNIGTVRTLRTKORIQPKTSYVI 180
Db 121 EVSIIOSMGYNRAKRLIOSEPENPSLOETSLVSQLSNIGTVRTLRTKORIQPKTSYVI 180
QY 181 ELGSDSSEDTVNKATYCSVGDDQLQITPQGRTRDEISLSDSAKKAACEFSETDVTNTEHQ 240
Db 181 ELGSDSSEDTVNKATYCSVGDDQLQITPQGRTRDEISLSDSAKKAACEFSETDVTNTEHQ 240
QY 241 PSNNDLNTTEKRAAERHPEKYQSSVSNSLHVPCGTNTHASSLOHENSLLTKTKORMNVE 300
Db 241 PSNNDLNTTEKRAAERHPEKYQSSVSNSLHVPCGTNTHASSLOHENSLLTKTKORMNVE 300
QY 301 KAEFCNKSQOPGLARSOHNRWAGSKETCNDRTPTSTKVKVDLNDAPLPCRKNKQKLP 360
Db 301 KAEFCNKSQOPGLARSOHNRWAGSKETCNDRTPTSTKVKVDLNDAPLPCRKNKQKLP 360
QY 361 SENPROTEDVPWITLNSSIOKYNEMFSRDELGGSDSDHSGESENNAKVADVDLNEVD 420
Db 361 SENPROTEDVPWITLNSSIOKYNEMFSRDELGGSDSDHSGESENNAKVADVDLNEVD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKKAASLPNLSHVTEN 480
Db 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKKAASLPNLSHVTEN 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFTKKADLAVOKTPEMINOQTNTE 540
Db 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFTKKADLAVOKTPEMINOQTNTE 540
QY 541 QNGQVNITNSGHENKTGDSIONEKNPNPIESLEKESAFKTKAEPISSSISNMELENI 600
Db 541 QNGQVNITNSGHENKTGDSIONEKNPNPIESLEKESAFKTKAEPISSSISNMELENI 600
QY 601 HNSKAPKNRLRKSSTRIHHALELWVSNLSPNCTELODSCSSSEIEIKKKYNQMPV 660
Db 601 HNSKAPKNRLRKSSTRIHHALELWVSNLSPNCTELODSCSSSEIEIKKKYNQMPV 660
QY 661 RHSRNLQMEGKEPATGAKSKNPNQETSKRHDSDTFPELKLTNAPGFTKCSNTSELKE 720
Db 661 RHSRNLQMEGKEPATGAKSKNPNQETSKRHDSDTFPELKLTNAPGFTKCSNTSELKE 720
QY 721 FVNPSLPREEKEKLETVKVSNAEDPKDMLSGSERVLTQERSVSSSISLVPGFDYGTQ 780
Db 721 FVNPSLPREEKEKLETVKVSNAEDPKDMLSGSERVLTQERSVSSSISLVPGFDYGTQ 780
QY 781 ESTSLLEVSTLGKAKTEPNKCVSOCAAFENPKGLIHGCSKDNNDTEGFKYPLGHEVNH 840
Db 781 ESTSLLEVSTLGKAKTEPNKCVSOCAAFENPKGLIHGCSKDNNDTEGFKYPLGHEVNH 840
QY 841 RETSIEMERSELDQYLONTFKVSKRQSPALFSPNGNAEEECATPSAHSGLSKQSPKVT 900
Db 841 RETSIEMERSELDQYLONTFKVSKRQSPALFSPNGNAEEECATPSAHSGLSKQSPKVT 900
QY 901 FECEQKEENOGKNESNIKPVQVNTITAGPVVQKDKPVDNAKCSIKGGRFCLSSQPRG 960
Db 901 FECEQKEENOGKNESNIKPVQVNTITAGPVVQKDKPVDNAKCSIKGGRFCLSSQPRG 960
QY 961 NETGLITPNKHGLQNPYRIPPLPIKSFVTKCKKNLLEENFESHSMSPREMGNEINIP 1020
Db 961 NETGLITPNKHGLQNPYRIPPLPIKSFVTKCKKNLLEENFESHSMSPREMGNEINIP 1020
QY 1021 STVSTISRNNIRENVFKEASSNINEVSSSTNEVGSSINEIGSSDENIQAEILGRNRGPKL 1080
Db 1021 STVSTISRNNIRENVFKEASSNINEVSSSTNEVGSSINEIGSSDENIQAEILGRNRGPKL 1080
QY 1081 NAMRLGVLOPEVYKOSLPGSNCKHPEIKKQBYEEVQVNTDFTSPYLLISDNLEQPMGSS 1140
Db 1081 NAMRLGVLOPEVYKOSLPGSNCKHPEIKKQBYEEVQVNTDFTSPYLLISDNLEQPMGSS 1140
QY 1141 HASQVCSETPDOLLDDGEIKEDTSAENDIKESSAVFSKSVQKGLSRSPFFTHLQ 1200
Db 1141 HASQVCSETPDOLLDDGEIKEDTSAENDIKESSAVFSKSVQKGLSRSPFFTHLQ 1200

QY 1201 GYRRGAKKLESSEENLSEDELPQFQHLFGKVNIPISQSTRHSTVATECLSKNTEENL 1260
Db 1201 GYRRGAKKLESSEENLSEDELPQFQHLFGKVNIPISQSTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDSCNQVILAKASQEHHLSEETKCSASFSSQCSSELEDLTANTNTQDPFLIGS 1320
Db 1261 LSLKNSLNDSCNQVILAKASQEHHLSEETKCSASFSSQCSSELEDLTANTNTQDPFLIGS 1320
QY 1321 SKQMRHQSESOGVGLSKDELVSDDERGTGLBENNQEQSMDSNLGEAASGCESETSVSE 1380
Db 1321 SKQMRHQSESOGVGLSKDELVSDDERGTGLBENNQEQSMDSNLGEAASGCESETSVSE 1380
QY 1381 DCSGLSSOSDILTTQORDTMQHNLIKLOEMAELAVLEQHGQSPSNSYPSIISDSSALE 1440
Db 1381 DCSGLSSOSDILTTQORDTMQHNLIKLOEMAELAVLEQHGQSPSNSYPSIISDSSALE 1440
QY 1441 DLRNPEQSTSEKAVLTSQKSSEYPISQNPEGLSADKFEVSADSSTSKNKEPGERSSPSK 1500
Db 1441 DLRNPEQSTSEKAVLTSQKSSEYPISQNPEGLSADKFEVSADSSTSKNKEPGERSSPSK 1500
QY 1501 CPSLDDRMYMHSCSGSLQNRNYPQOEELIKVYDVVEQOOLEESGPHDLTETSYPQLDLEG 1560
Db 1501 CPSLDDRMYMHSCSGSLQNRNYPQOEELIKVYDVVEQOOLEESGPHDLTETSYPQLDLEG 1560
QY 1561 TPYLESGLISLSDPESDPSEDRAPEARVGNIPSTTSALKVPQLKVAESAQSPAAHHT 1620
Db 1561 TPYLESGLISLSDPESDPSEDRAPEARVGNIPSTTSALKVPQLKVAESAQSPAAHHT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERVNKRMSMVVSGLTPEEPMLVYKFARKHHTLTNLI 1680
Db 1621 DTAGYNAMESVSREKPELTASTERVNKRMSMVVSGLTPEEPMLVYKFARKHHTLTNLI 1680
QY 1681 TEETHVVMKTDAEVCEBTLKYLFIAGGKVVVYFWVTQSIKERKMLNEHDFEVRGDV 1740
Db 1681 TEETHVVMKTDAEVCEBTLKYLFIAGGKVVVYFWVTQSIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHQGPKRARESDRKIFRGLIEICCYGPTNPTDQLEMMVOLCGASVVKELSSFTL 1800
Db 1741 VNGRNHQGPKRARESDRKIFRGLIEICCYGPTNPTDQLEMMVOLCGASVVKELSSFTL 1800
QY 1801 GTGVHPIVVQPDWATEDNGFHAIGOMCEAPVVTREWVLDVSVALYQCOELDTYLIPOIPH 1860
Db 1801 GTGVHPIVVQPDWATEDNGFHAIGOMCEAPVVTREWVLDVSVALYQCOELDTYLIPOIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863

RESULT 5

US-08-483-553-2
; Sequence 2, Application US/08483553
; Patent No. 5709999
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA

Db 1501 CPSLDRWMHSCSGSLQNRNYPSEELIKVVDVEEQOLESGPHDLTETSYLPQDLEG 1560
Qy 1561 TPYLESGISLFDSPEDSPEDRAPESARVGNIPSTTSALKVPOLKVAESAQSPAAAHTT 1620
Db 1561 TPYLESGISLFDSPEDSPEDRAPESARVGNIPSTTSALKVPOLKVAESAQSPAAAHTT 1620
Qy 1621 DTAGYNAMESYSREKPELTASTERVNKRMSVMVYSGLTPEEFMLVYKFARKHHITLTNLI 1680
Db 1621 DTAGYNAMESYSREKPELTASTERVNKRMSVMVYSGLTPEEFMLVYKFARKHHITLTNLI 1680
Qy 1681 TEETHVVMKTDAEVFCERTLKYFLGIAGGKWVYSYFWVTOSIKERKMLNHDHFEVRGDV 1740
Db 1681 TEETHVVMKTDAEVFCERTLKYFLGIAGGKWVYSYFWVTOSIKERKMLNHDHFEVRGDV 1740
Qy 1741 VNGRNHQGPKRARESDQRKIRFGLIEICCYGPTNNMPTDQLEMMVOLCGASVVKLSFTL 1800
Db 1741 VNGRNHQGPKRARESDQRKIRFGLIEICCYGPTNNMPTDQLEMMVOLCGASVVKLSFTL 1800
Qy 1801 GTGVHPVVVQPDAWTNGFHAIGQMCEAPVVTREWVLDVSVALYQCELDTYLIPQIPH 1860
Db 1801 GTGVHPVVVQPDAWTNGFHAIGQMCEAPVVTREWVLDVSVALYQCELDTYLIPQIPH 1860
Qy 1861 SHY 1863
Db 1861 SHY 1863

RESULT 6
US-08-487-002-2
; Sequence 2, Application US/08487002
; Patent No. 5710001
; GENERAL INFORMATION:
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Simard, Jacques
; APPLICANT: Emi, Mitsuru
; APPLICANT: Nakamura, Yusuke
; APPLICANT: Durocher, Francine
; TITLE OF INVENTION: 17q-linked Breast and Ovarian Cancer
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,002
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.

; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-487-002-2

Query Match 99.9%; Score 9642; DB 1; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MDLSARVEEVQNVINAMQKILECPICLELIKPEVSTKCDHIFCKFCMLKLLNKKGPSQ 60
Db 1 MDLSARVEEVQNVINAMQKILECPICLELIKPEVSTKCDHIFCKFCMLKLLNKKGPSQ 60
Qy 61 CPICKNDITKRSLOESTRFSQLVVEELKIIICAFQLDTGLEYANSYNFAKKENNSPEHLKD 120
Db 61 CPICKNDITKRSLOESTRFSQLVVEELKIIICAFQLDTGLEYANSYNFAKKENNSPEHLKD 120
Qy 121 EVSIIQSMGYNRNRAKRLLOQSEPNPSLOETSLVSQLSNLGTVRTLTQRIQPOKTSYVI 180
Db 121 EVSIIQSMGYNRNRAKRLLOQSEPNPSLOETSLVSQLSNLGTVRTLTQRIQPOKTSYVI 180
Qy 181 ELGSDSSEDVTNKATYCSVGDDQELLQITPQGTTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDVTNKATYCSVGDDQELLQITPQGTTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Qy 241 PSNNDLNTTEKRAAERPEKYQGSVSNLHVEPCGTNTHASSLOHENSLLLTKDRMNV 300
Db 241 PSNNDLNTTEKRAAERPEKYQGSVSNLHVEPCGTNTHASSLOHENSLLLTKDRMNV 300
Qy 301 KAECFNKSKOPGLARSOHNRWAGSKETCNDRTPTSTKPKVDLNADPLCERKEWNKKQLPC 360
Db 301 KAECFNKSKOPGLARSOHNRWAGSKETCNDRTPTSTKPKVDLNADPLCERKEWNKKQLPC 360
Qy 361 SENPRDTEVPWITLNSSIQKVNEWFSRSDDELGSDSDHGESESNKAVADVLDVNEVD 420
Db 361 SENPRDTEVPWITLNSSIQKVNEWFSRSDDELGSDSDHGESESESNKAVADVLDVNEVD 420
Qy 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKASLPNLSHVTE 480
Db 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKASLPNLSHVTE 480
Qy 481 LIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPEDFTKKADLAVQKTPMINOCTNOTE 540
Db 481 LIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPEDFTKKADLAVQKTPMINOCTNOTE 540
Qy 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPTLESLEKESAFTKAEPTSSSISSNMELELNI 600
Db 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPTLESLEKESAFTKAEPTSSSISSNMELELNI 600
Qy 601 HNSKAPKKNRLRRKSSRTHIALELVYSRNLSPNCTELQIDSCSSSEIEIKKKYQNPV 660
Db 601 HNSKAPKKNRLRRKSSRTHIALELVYSRNLSPNCTELQIDSCSSSEIEIKKKYQNPV 660
Qy 661 RHRNRLQMEGKEPATGAKKSNKNEOTSKRHSDTPELKLNTNAPGSFTKCSNTSELKE 720
Db 661 RHRNRLQMEGKEPATGAKKSNKNEOTSKRHSDTPELKLNTNAPGSFTKCSNTSELKE 720
Qy 721 FVNPSPREEKEKLETYKVSNNADPKDMLSGERVLTQTSRVSESSISLVPQTDYGTQ 780
Db 721 FVNPSPREEKEKLETYKVSNNADPKDMLSGERVLTQTSRVSESSISLVPQTDYGTQ 780
Qy 781 ESISLLEVSTLGRKAKTEPNKCVSOCAAFENPKGLIHGCSKDRNNDTEGFKYPLGHEVNH 840
Db 781 ESISLLEVSTLGRKAKTEPNKCVSOCAAFENPKGLIHGCSKDRNNDTEGFKYPLGHEVNH 840

QY 841 RETSIEMESELDAQYLQNTFKVKSKROSFALFSPNGNAEEBCATFSAHSGSLKKQSPKVT 900
Db 841 RETSIEMESELDAQYLQNTFKVKSKROSFALFSPNGNAEEBCATFSAHSGSLKKQSPKVT 900
QY 901 FECQKKEQKKNESNFKPVQTVNITAGFPVVGQKDPVDNAKCSIKGGRFCLSSQFRG 960
Db 901 FECQKKEQKKNESNFKPVQTVNITAGFPVVGQKDPVDNAKCSIKGGRFCLSSQFRG 960
QY 961 NETGLITPNKHGGLQNPYRPPLEPKISFVKTKCKKLLLENFEEHSMSPEREMGNENIP 1020
Db 961 NETGLITPNKHGGLQNPYRPPLEPKISFVKTKCKKLLLENFEEHSMSPEREMGNENIP 1020
QY 1021 STVTISRNNIRENVFKEASSNINEYSSSTNEVGSSSINEIGSSDENIQAEIGRNRGPKL 1080
Db 1021 STVTISRNNIRENVFKEASSNINEYSSSTNEVGSSSINEIGSSDENIQAEIGRNRGPKL 1080
QY 1081 NAMLRGLVQPEVYKQSLPGSNCKHPEIKKOEYEEVQTVNTDPSPLYISDNLEQPMGSS 1140
Db 1081 NAMLRGLVQPEVYKQSLPGSNCKHPEIKKOEYEEVQTVNTDPSPLYISDNLEQPMGSS 1140
QY 1141 HASQVCSETPDDLLDGEIKEDTSAFENDIKESSAVFSKSVQKGLSRSPFTHTHLAQ 1200
Db 1141 HASQVCSETPDDLLDGEIKEDTSAFENDIKESSAVFSKSVQKGLSRSPFTHTHLAQ 1200
QY 1201 GYRRGAKKLESSEENLSEDEELPCFQHLFGKVNIPSTQSTHSTVATECLSKNTEENL 1260
Db 1201 GYRRGAKKLESSEENLSEDEELPCFQHLFGKVNIPSTQSTHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLDCSNQVILAKAQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS 1320
Db 1261 LSLKNSLDCSNQVILAKAQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS 1320
QY 1321 SKQMRHOSQGVGLSKELVSDDEERGTLGLENNQEQSMDSNLGEAAGCESETSVSE 1380
Db 1321 SKQMRHOSQGVGLSKELVSDDEERGTLGLENNQEQSMDSNLGEAAGCESETSVSE 1380
QY 1381 DCSGLSSQSDILTTQQRDTMOHNLKIQEAMAEVLEQHGQSPNSYPSIISDSSALE 1440
Db 1381 DCSGLSSQSDILTTQQRDTMOHNLKIQEAMAEVLEQHGQSPNSYPSIISDSSALE 1440
QY 1441 DLRNPEOSTSEKAVLTQKSEYPISONPEGLSADKEFEVSADSTSKNKEPGEVRSRSPSK 1500
Db 1441 DLRNPEOSTSEKAVLTQKSEYPISONPEGLSADKEFEVSADSTSKNKEPGEVRSRSPSK 1500
QY 1501 CPSLDDRWYMHSCGSLQNRNYPQOEELIKVDVVEEQOLEESGPHDLTETSYLPRQDLEG 1560
Db 1501 CPSLDDRWYMHSCGSLQNRNYPQOEELIKVDVVEEQOLEESGPHDLTETSYLPRQDLEG 1560
QY 1561 TPYLESGISLFSDDPESDRAPEARSVGNIPSTTSALKVPOLKVAESAQAASAAHTT 1620
Db 1561 TPYLESGISLFSDDPESDRAPEARSVGNIPSTTSALKVPOLKVAESAQAASAAHTT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERNKRMVMVSGLTPEEFMLVYKFAKHHITLNL 1680
Db 1621 DTAGYNAMESVSREKPELTASTERNKRMVMVSGLTPEEFMLVYKFAKHHITLNL 1680
QY 1681 TEETHVMKTDAEFVCERTILKYFLGAGGKVVYVFWVTQSIKERKMLNEHDEFEVGDV 1740
Db 1681 TEETHVMKTDAEFVCERTILKYFLGAGGKVVYVFWVTQSIKERKMLNEHDEFEVGDV 1740
QY 1741 VNGRNHQGPKRARSQDRKIFRGLIEICCYGPFNTMPDQLEWMVQLCGASVYKELSSFTL 1800
Db 1741 VNGRNHQGPKRARSQDRKIFRGLIEICCYGPFNTMPDQLEWMVQLCGASVYKELSSFTL 1800
QY 1801 GTGVHPITVVVQPDAAWEDNGFHAIGQCEAPVVTREWVLDSVALYQOEQLDTYLIPQIPH 1860
Db 1801 GTGVHPITVVVQPDAAWEDNGFHAIGQCEAPVVTREWVLDSVALYQOEQLDTYLIPQIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863

RESULT 7
US-08-483-554B-2
: Sequence 2, Application US/08483554B
: Patent No. 5747282
: GENERAL INFORMATION:
: APPLICANT: Skolnick, Mark H.
: APPLICANT: Goldgar, David E.
: APPLICANT: Miki, Yoshio
: APPLICANT: Swenson, Jeff
: APPLICANT: Kamb, Alexander
: APPLICANT: Harshman, Keith D.
: APPLICANT: Shattuck-Eidens, Donna M.
: APPLICANT: Tavtigian, Sean V.
: APPLICANT: Wiseman, Roger W.
: APPLICANT: Futreal, P. Andrew
: TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
: TITLE OF INVENTION: Susceptibility Gene
: NUMBER OF SEQUENCES: 85
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
: STREET: 1201 New York Avenue, N.W., Suite 1000
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/483,554B
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/409,305
: FILING DATE: 24-MAR-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/348,824
: FILING DATE: 29-NOV-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/308,104
: FILING DATE: 16-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/300,266
: FILING DATE: 02-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/289,221
: FILING DATE: 12-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Ihnen, Jeffrey L.
: REGISTRATION NUMBER: 28,957
: REFERENCE/DOCKET NUMBER: 24884-109347
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-962-4810
: TELEFAX: 202-962-8300
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1863 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-483-554B-2

Query Match 99.9%; Score 9642; DB 1; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDLSALRVEEVQNVINAMQKILEPICLELTKEPVSTKCDHIECKFCMLKLLNOKKGPSQ 60
Db 1 MDLSALRVEEVQNVINAMQKILEPICLELTKEPVSTKCDHIECKFCMLKLLNOKKGPSQ 60
QY 61 CPLCKNDITKRSIQESTRFSQLVVEELLKIICAFOLDGTGLEVANSYNFAKKENNSPEHLKD 120

Db 61 CPCLCKNDITKRSQESTRFSQVVEELKIIICAFQDITGLEVANSYNFAKKENNSPEHLKD 120
QY 121 EVSIIOSMGVNRKAKLLQSEPNPSLQETSLSVQLSNLGTVRTLRTKRIOPQKTSYVI 180
Db 121 EVSIIOSMGVNRKAKLLQSEPNPSLQETSLSVQLSNLGTVRTLRTKRIOPQKTSYVI 180
QY 181 ELGSDSSEDTVNKATYCSVGQDQELQITPQGTTRDEISLSDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDTVNKATYCSVGQDQELQITPQGTTRDEISLSDSAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNNDLNTTEKRAAERHPKPYOGSSVSNLHVPCGTNTNTHASSLQENSHLLLTJKDRMNV 300
Db 241 PSNNDLNTTEKRAAERHPKPYOGSSVSNLHVPCGTNTNTHASSLQENSHLLLTJKDRMNV 300
QY 301 KAFECFKSKOPGLARSOHRNWRAGSKETCNDRTPTSTKKYVDLNADPLCKERWNKQKLP 360
Db 301 KAFECFKSKOPGLARSOHRNWRAGSKETCNDRTPTSTKKYVDLNADPLCKERWNKQKLP 360
QY 361 SENPRDTEVPWITLNSSTIQKYNWFNRSDDELIGSDSDHGESESNKAVADVLDVINEVD 420
Db 361 SENPRDTEVPWITLNSSTIQKYNWFNRSDDELIGSDSDHGESESNKAVADVLDVINEVD 420
QY 421 EYSGSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKRAKASLPNLSHVTE 480
Db 421 EYSGSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKRAKASLPNLSHVTE 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRRTPSGLHPEDFIKKADLAVOKTPEMINOQTNTE 540
Db 481 LIIGAFVTEPQIIQERPLTNKLRKRRTPSGLHPEDFIKKADLAVOKTPEMINOQTNTE 540
QY 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNMELEINI 600
Db 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNMELEINI 600
QY 601 HNSKAPKKNLRKRSSTRIHAELELVVSRNLSPNCTELQIDSCSSSEIEIKKKYNQMPV 660
Db 601 HNSKAPKKNLRKRSSTRIHAELELVVSRNLSPNCTELQIDSCSSSEIEIKKKYNQMPV 660
QY 661 RHNRNLQMEGEPATGAKSKNPQETSKRHDSDTFPELKLITNAPGFTKCSNTSELKE 720
Db 661 RHNRNLQMEGEPATGAKSKNPQETSKRHDSDTFPELKLITNAPGFTKCSNTSELKE 720
QY 721 FVNPSLPREEKEKLETVKVSNAEDPKDMLSGERVLTQTERSVESSISLVPGTDYGTQ 780
Db 721 FVNPSLPREEKEKLETVKVSNAEDPKDMLSGERVLTQTERSVESSISLVPGTDYGTQ 780
QY 781 ESISLLEVSTLGAKTEPNKCVSOCAAFENPKGLIHGCSKONRNDTEGFKYPLGHEVNH 840
Db 781 ESISLLEVSTLGAKTEPNKCVSOCAAFENPKGLIHGCSKONRNDTEGFKYPLGHEVNH 840
QY 841 RETSIEMEESLDAQYLONTFKVSKRQSFALPSNPGNAEEECATSAHSGSLKKOSPKVT 900
Db 841 RETSIEMEESLDAQYLONTFKVSKRQSFALPSNPGNAEEECATSAHSGSLKKOSPKVT 900
QY 901 FECEQKEENQGNESNIKPVQVTNITAGFPVVGQDKDPVDNAKCSIKGSRFCLSSQFRG 960
Db 901 FECEQKEENQGNESNIKPVQVTNITAGFPVVGQDKDPVDNAKCSIKGSRFCLSSQFRG 960
QY 961 NETGLITPNKHGLLQNPYRIPPLPIKGFVKTKCKNLLLENFEBHSMSPREMGNEINP 1020
Db 961 NETGLITPNKHGLLQNPYRIPPLPIKGFVKTKCKNLLLENFEBHSMSPREMGNEINP 1020
QY 1021 STVSTISRNRIENVFKEASSNINEVGSSTNEVGSSINEIGSSDENIQAEIAGNRGPKL 1080
Db 1021 STVSTISRNRIENVFKEASSNINEVGSSTNEVGSSINEIGSSDENIQAEIAGNRGPKL 1080
QY 1081 NAMRLGVLOPEVYKQSLPGSKCKPEIKKQYEEVQVTNTDFSPYILISNLEQPMGSS 1140
Db 1081 NAMRLGVLOPEVYKQSLPGSKCKPEIKKQYEEVQVTNTDFSPYILISNLEQPMGSS 1140
QY 1141 HASQVCSETPDLDLDDGEIKEDTSAENDIKESSAVFSKSVQKGELSRSPPFTHTHLAQ 1200
|||||

Db 1141 HASQVCSETPDLDLDDGEIKEDTSAENDIKESSAVFSKSVQKGELSRSPPFTHTHLAQ 1200
QY 1201 GYRRGAKKLESSEENLSSEDEELPCFOHLLFKGVNNIPSQSTRHSHTVATECLSKNTEENL 1260
Db 1201 GYRRGAKKLESSEENLSSEDEELPCFOHLLFKGVNNIPSQSTRHSHTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDSCSNQVITLAKASQEHHLSEETKCSASFSSQCSSELEDLTANTNTQDPPLIGS 1320
Db 1261 LSLKNSLNDSCSNQVITLAKASQEHHLSEETKCSASFSSQCSSELEDLTANTNTQDPPLIGS 1320
QY 1321 SKQMRHQSESQGVGLSKDELVSDDERGTGLEENNOEBOQSMDSNLGEEASGESETSVSE 1380
Db 1321 SKQMRHQSESQGVGLSKDELVSDDERGTGLEENNOEBOQSMDSNLGEEASGESETSVSE 1380
QY 1381 DCSGLSSQSDILTTQORDTMQHNLIKLOEMAELEAVLEOHGSQPSNSYPSIISDSALE 1440
Db 1381 DCSGLSSQSDILTTQORDTMQHNLIKLOEMAELEAVLEOHGSQPSNSYPSIISDSALE 1440
QY 1441 DLNRNPEQSTSEKAVLTSQKSEYPISONPEGLSADKFEVSADSSSTSKNKEPGVERSPSPK 1500
Db 1441 DLNRNPEQSTSEKAVLTSQKSEYPISONPEGLSADKFEVSADSSSTSKNKEPGVERSPSPK 1500
QY 1501 CPSLDDRWMHSCSGSLQNRNYPQSEELIKVVDVBEQQLEESGPHDLTETSYLPRODLEG 1560
Db 1501 CPSLDDRWMHSCSGSLQNRNYPQSEELIKVVDVBEQQLEESGPHDLTETSYLPRODLEG 1560
QY 1561 TPYLESGISLSDPESDPEDRAPESARVGNIPSTTSALKVPOLKVAESAQSPAAHHT 1620
Db 1561 TPYLESGISLSDPESDPEDRAPESARVGNIPSTTSALKVPOLKVAESAQSPAAHHT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERVNKRMSMVVSGLTPPEEFMLVYKFARKHHITLTNLI 1680
Db 1621 DTAGYNAMESVSREKPELTASTERVNKRMSMVVSGLTPPEEFMLVYKFARKHHITLTNLI 1680
QY 1681 TEETHVVMKTDAEFVCERTLYKFLGIAGGKVVVSYFWVTQSIKERKMLNEHDFEVRGDV 1740
Db 1681 TEETHVVMKTDAEFVCERTLYKFLGIAGGKVVVSYFWVTQSIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHQPKRAESQDRKIPRGLEICCYGPTNPTDOLEHVMVOLCGASVVKELSSFTL 1800
Db 1741 VNGRNHQPKRAESQDRKIPRGLEICCYGPTNPTDOLEHVMVOLCGASVVKELSSFTL 1800
QY 1801 GTGVHPIVVVPDAMTNGFHAIQMCEAPVVTREWLDSVALYQCQELDLYLIPQIPH 1860
Db 1801 GTGVHPIVVVPDAMTNGFHAIQMCEAPVVTREWLDSVALYQCQELDLYLIPQIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863
RESULT 8
US-08-488-011B-2
; Sequence 2, Application US/08488011B
; Patent No. 5753441
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington

STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,011B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Innen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347-09
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-488-011B-2

Query Match 99.9%; Score 9642; DB 1; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	MDLSALRVEEVQNVINAMQKILECPICLLEIKPEVSTKCDHIFCKFCMCLKLLNQKKGPSQ	60
Db	1	MDLSALRVEEVQNVINAMQKILECPICLLEIKPEVSTKCDHIFCKFCMCLKLLNQKKGPSQ	60
QY	61	CPLCKNDITKRSLOESTRFSQVVEELKIIICAFQLDITGLEAYANSYFNAKKENNSPEHLKD	120
Db	61	CPLCKNDITKRSLOESTRFSQVVEELKIIICAFQLDITGLEAYANSYFNAKKENNSPEHLKD	120
QY	121	EVSIIQSMGYRNRAKRLLOSEPNSLOETSLSVQLSNLGTVRTLRKQRIQPKTSVYI	180
Db	121	EVSIIQSMGYRNRAKRLLOSEPNSLOETSLSVQLSNLGTVRTLRKQRIQPKTSVYI	180
QY	181	ELGSDSSEDTVNKATYCSVGDQELLIQTQGTDRDEISLSAKKAACEFSETDVTNTEHQ	240
Db	181	ELGSDSSEDTVNKATYCSVGDQELLIQTQGTDRDEISLSAKKAACEFSETDVTNTEHQ	240
QY	241	PSNNDLNTTEKRAAEHHPKYGSSVSNLHVPCGNTTHASSLOHENSLLLTDKRMNVE	300
Db	241	PSNNDLNTTEKRAAEHHPKYGSSVSNLHVPCGNTTHASSLOHENSLLLTDKRMNVE	300
QY	301	KAIEFCNKSQKQGLARSQHNWAGSKETCNDNDRTPSTEKKVDLNADPLCERKEWNKOKLPC	360
Db	301	KAIEFCNKSQKQGLARSQHNWAGSKETCNDNDRTPSTEKKVDLNADPLCERKEWNKOKLPC	360
QY	361	SENPRDTEVPWITLNSITQKYNEWFRSDELLGSDSDHSGESENKAYVDVLDVLYNEVD	420
Db	361	SENPRDTEVPWITLNSITQKYNEWFRSDELLGSDSDHSGESENKAYVDVLDVLYNEVD	420

QY	421	EYSGSSEKIDLLASDPHEALICKSERVHSHKSVESNIETDKIFGKTYRKKASLPNLSHVTEN	480
Db	421	EYSGSSEKIDLLASDPHEALICKSERVHSHKSVESNIETDKIFGKTYRKKASLPNLSHVTEN	480
QY	481	LIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPEDFIKKADLAVQKTPEMINOQNTQTE	540
Db	481	LIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPEDFIKKADLAVQKTPEMINOQNTQTE	540
QY	541	QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINWLELNI	600
Db	541	QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINWLELNI	600
QY	601	HNSKAPKKNRLRRKSSTRIHALELVVSRNLSPPNCTELQIDSCSSSEBIKKKKYNQMPV	660
Db	601	HNSKAPKKNRLRRKSSTRIHALELVVSRNLSPPNCTELQIDSCSSSEBIKKKKYNQMPV	660
QY	661	RHSRNLQMEGKEPATGAKKSNKPNEQTSKRHDSDFPELKTNPAGSFTKCSNTSELKE	720
Db	661	RHSRNLQMEGKEPATGAKKSNKPNEQTSKRHDSDFPELKTNPAGSFTKCSNTSELKE	720
QY	721	FYNPSLPREEKEELKLVKVSNNADPKDMLSGERVLOTERSVESSSISLVPGTDYGTQ	780
Db	721	FYNPSLPREEKEELKLVKVSNNADPKDMLSGERVLOTERSVESSSISLVPGTDYGTQ	780
QY	781	ESISLLEVTSLGKAKTEPNKCVSQAAPENPKGLIHGCSKDNRNDETEGKYPLGHEVNH	840
Db	781	ESISLLEVTSLGKAKTEPNKCVSQAAPENPKGLIHGCSKDNRNDETEGKYPLGHEVNH	840
QY	841	RETSIEMESELDAQYLQNTFKVSKRQSPALFSNPGNAEEECATPSAHSGLKQSPKVT	900
Db	841	RETSIEMESELDAQYLQNTFKVSKRQSPALFSNPGNAEEECATPSAHSGLKQSPKVT	900
QY	901	FECEQKEENQGNESNIKPVQTVNITAGFPVVGQKDKPVDNAKCSIKGSRCLSSQFRG	960
Db	901	FECEQKEENQGNESNIKPVQTVNITAGFPVVGQKDKPVDNAKCSIKGSRCLSSQFRG	960
QY	961	NETGLITPNKHGLLQNPYRIPPLPIKSFVKTKCKNLLLENFEEHSMSPERMGNENIP	1020
Db	961	NETGLITPNKHGLLQNPYRIPPLPIKSFVKTKCKNLLLENFEEHSMSPERMGNENIP	1020
QY	1021	STVSTISRNNIRENVFKEASSNINEVGSTNEVGSSSINEIGSSDENIOAELGRNRPKL	1080
Db	1021	STVSTISRNNIRENVFKEASSNINEVGSTNEVGSSSINEIGSSDENIOAELGRNRPKL	1080
QY	1081	NAMRLGVLOPEVYKQSLPGSNCKHPEIKKQYEEVQVTVNTDFSPYLISDNLEQPMGSS	1140
Db	1081	NAMRLGVLOPEVYKQSLPGSNCKHPEIKKQYEEVQVTVNTDFSPYLISDNLEQPMGSS	1140
QY	1141	HASQVCSETPDDLLDDGEIKEDTSAENDIKESSAVFSKSVOKGELSRSPSPFTHTHQAQ	1200
Db	1141	HASQVCSETPDDLLDDGEIKEDTSAENDIKESSAVFSKSVOKGELSRSPSPFTHTHQAQ	1200
QY	1201	GYRRGAKKLESSEENLSSDEELPCFQHLFGKVNINIPSQSTRHSTVATECLSKNTEENL	1260
Db	1201	GYRRGAKKLESSEENLSSDEELPCFQHLFGKVNINIPSQSTRHSTVATECLSKNTEENL	1260
QY	1261	LSLKNSLNDCSNQVILAKASQEHHLSEETKCSASFSSOCSELEDLTANTNTQDPFLIGS	1320
Db	1261	LSLKNSLNDCSNQVILAKASQEHHLSEETKCSASFSSOCSELEDLTANTNTQDPFLIGS	1320
QY	1321	SKQMRHOSQGVGLSDKELVSDDEERGTLLENNOEESMDSNLGEAAGCESETSUSE	1380
Db	1321	SKQMRHOSQGVGLSDKELVSDDEERGTLLENNOEESMDSNLGEAAGCESETSUSE	1380
QY	1381	DCSGLSSQSDILTTQORDTMQHNLKIQOEMAEALVLEQHGHSQSPNSYPSIISDSSALE	1440
Db	1381	DCSGLSSQSDILTTQORDTMQHNLKIQOEMAEALVLEQHGHSQSPNSYPSIISDSSALE	1440
QY	1441	DLRNPQSTSEKAVLTQSQSEYPIQSNPEGLSADKFEVSADSSTSKNKEPQGVRSRSPSK	1500
Db	1441	DLRNPQSTSEKAVLTQSQSEYPIQSNPEGLSADKFEVSADSSTSKNKEPQGVRSRSPSK	1500

QY 1501 CPSLDRWYMHSCGSLQNRNYPQSEELIKVVVDVEEQLEESGPHDLTETSYLPRQDLEG 1560
Db 1501 CPSLDRWYMHSCGSLQNRNYPQSEELIKVVVDVEEQLEESGPHDLTETSYLPRQDLEG 1560
QY 1561 TPLYSGISLFDSPEDSPEDRAPESARVGNIPSTTSALKVPQKVAESAQSPAAAHHT 1620
Db 1561 TPLYSGISLFDSPEDSPEDRAPESARVGNIPSTTSALKVPQKVAESAQSPAAAHHT 1620
QY 1621 DTAGYNAMESVSRKPELTASTERVNRKMSVWVSGLTPEEFMLVYKPAKHHITLTNLI 1680
Db 1621 DTAGYNAMESVSRKPELTASTERVNRKMSVWVSGLTPEEFMLVYKPAKHHITLTNLI 1680
QY 1681 TEETHVVMKTDAEVCERTLKIFLGIAGGKVVSYFWVTQSIKIRKMLNHDHFEVRGDV 1740
Db 1681 TEETHVVMKTDAEVCERTLKIFLGIAGGKVVSYFWVTQSIKIRKMLNHDHFEVRGDV 1740
QY 1741 VNGRNHOGPKRAESQDRKIFRGLBICCYGPTNNPTDQLEWMVOLCGASVVKELSSFTL 1800
Db 1741 VNGRNHOGPKRAESQDRKIFRGLBICCYGPTNNPTDQLEWMVOLCGASVVKELSSFTL 1800
QY 1801 GTGVHPVVVQPDWNTDNGFHAIGOMCEAPVVTREWVLDSDVALYQCOELDTYLIPOLPH 1860
Db 1801 GTGVHPVVVQPDWNTDNGFHAIGOMCEAPVVTREWVLDSDVALYQCOELDTYLIPOLPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863
RESULT 9
US-08-850-727-2
; Sequence 2, Application US/08850727
; Patent No. 6162897
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/850,727
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/483,554
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-850-727-2
Query Match 99.98; Score 9642; DB 4; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MDLSARVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNOKKGPQ 60
Db 1 MDLSARVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNOKKGPQ 60
QY 61 CPLCKNDITKRSLOESTRFSQVLSVEELLKIIICAFOLDTGLEVANSYNFAKKNNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRFSQVLSVEELLKIIICAFOLDTGLEVANSYNFAKKNNSPEHLKD 120
QY 121 EVSIIQSMGYNRNRAKRLLOSEPENPSLQETSLVQLSLNLTGTVRLTKQRTQPKTSYVI 180
Db 121 EVSIIQSMGYNRNRAKRLLOSEPENPSLQETSLVQLSLNLTGTVRLTKQRTQPKTSYVI 180
QY 181 ELGSDSSEDVTNKATYCSVGDOELLQITPQGTDRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDVTNKATYCSVGDOELLQITPQGTDRDEISLDSAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNNDLNTTEKRAAERPEKYQGVSSVNLHVPCGTNTHASSLQHENSLLLTQDRMVE 300
Db 241 PSNNDLNTTEKRAAERPEKYQGVSSVNLHVPCGTNTHASSLQHENSLLLTQDRMVE 300
QY 301 KAEPCNKSQOPGLARSOHNRWAGSKETCNDRTPTSTKVKVDLNDPLCERKEWNNKQKLP 360
Db 301 KAEPCNKSQOPGLARSOHNRWAGSKETCNDRTPTSTKVKVDLNDPLCERKEWNNKQKLP 360
QY 361 SENPRDTEVPWITLNSSIQKVNWFSSRDELGSDSDSHDGESESNKAVADVLDVINEVD 420
Db 361 SENPRDTEVPWITLNSSIQKVNWFSSRDELGSDSDSHDGESESNKAVADVLDVINEVD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKYRKKASLPNLSHVTE 480
Db 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKYRKKASLPNLSHVTE 480
QY 481 LIICAFVTEPQIIQERPLTNKLRKRPTSGLHPEDFTKKADLAVOKTPEMINOCTNOTE 540
Db 481 LIICAFVTEPQIIQERPLTNKLRKRPTSGLHPEDFTKKADLAVOKTPEMINOCTNOTE 540
QY 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPTESLEKESAFKTKAEPISSSISNMELELNI 600
Db 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPTESLEKESAFKTKAEPISSSISNMELELNI 600
QY 601 HNSKAPKKNLRRKSSRTHIALELVYSRNLSPNCTQELQIDSCSSSEIEIKKKYQNPV 660
Db 601 HNSKAPKKNLRRKSSRTHIALELVYSRNLSPNCTQELQIDSCSSSEIEIKKKYQNPV 660
QY 661 RHRNRLQMEGKEPATGAKKSNKNEQTSKRHSDTPELKLTNAPGSFTKCSNTSELKE 720
Db 661 RHRNRLQMEGKEPATGAKKSNKNEQTSKRHSDTPELKLTNAPGSFTKCSNTSELKE 720
QY 721 FVNPSPREEKEKLETYKVSNNAEKPKDMLSLGERSVLSVPGTDYGTQ 780

Db 721 FVNPSPREEKEKLETVKYSVNAEDPKDMLSGSERVLQTERSVESSISLVPDYGTO 780
QY 781 ESISLLEVTSLGKAKTEPNKCSQCAAFENPKGLIHGCCSKDNENDTEGPKYPLGHEVNH 840
Db 781 ESISLLEVTSLGKAKTEPNKCSQCAAFENPKGLIHGCCSKDNNDTEGPKYPLGHEVNH 840
QY 841 RETSIEMESELDAOYLQNTFKYVKRQSFALFSPNCAEBECATFSAHSGSLKKQSPKVT 900
Db 841 RETSIEMESELDAOYLQNTFKYVKRQSFALFSPNCAEBECATFSAHSGSLKKQSPKVT 900
QY 901 FECEQKEENOGKNEISKPVQTVNITAGPVPVQKDKPDVNAKCSIKGGSRFCLSQFRG 960
Db 901 FECEQKEENOGKNEISKPVQTVNITAGPVPVQKDKPDVNAKCSIKGGSRFCLSQFRG 960
QY 961 NETGLTTPKNHGLLQNPYRTPPLFPKTSVKTKCKNLLLEENFEHSMSPEREMGNENIP 1020
Db 961 NETGLTTPKNHGLLQNPYRTPPLFPKTSVKTKCKNLLLEENFEHSMSPEREMGNENIP 1020
QY 1021 STVSTISRNNIRENVFKEASSNINEVGSTNEVGSSINEIGSSDENIOAELGRNRPKL 1080
Db 1021 STVSTISRNNIRENVFKEASSNINEVGSTNEVGSSINEIGSSDENIOAELGRNRPKL 1080
QY 1081 NAMRLGVLOPEVYKOSLPGSNCKHPEIKKQYEEVVQTVNTDFSPYLSIDNLEQPMGSS 1140
Db 1081 NAMRLGVLOPEVYKOSLPGSNCKHPEIKKQYEEVVQTVNTDFSPYLSIDNLEQPMGSS 1140
QY 1141 HASOVCSPTDDLDDGETIKEDTSPAENDIKESSAVFSKVGKELSRSPSPFTHHLAQ 1200
Db 1141 HASOVCSPTDDLDDGETIKEDTSPAENDIKESSAVFSKVGKELSRSPSPFTHHLAQ 1200
QY 1201 GYRRGAKKLESSEENLSSEDEELPCFQHLFGKVNIPQSQRHSTVATECLSKNTEENL 1260
Db 1201 GYRRGAKKLESSEENLSSEDEELPCFQHLFGKVNIPQSQRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLDCSNQVILAKASQEHLSSEETKCSASLFSQCSSELEDLTANTNTQDPFLIG 1320
Db 1261 LSLKNSLDCSNQVILAKASQEHLSSEETKCSASLFSQCSSELEDLTANTNTQDPFLIG 1320
QY 1321 SKOMRHQESQGVGLSDKELVDDEERGTLGLENNOEQSMDSNLGEASGCESETSVSE 1380
Db 1321 SKOMRHQESQGVGLSDKELVDDEERGTLGLENNOEQSMDSNLGEASGCESETSVSE 1380
QY 1381 DCSGLSSQSDILTTQQRDTMQHNLKIQOEMAELEAVLHQHQSQPSNSYPSIISDSSALE 1440
Db 1381 DCSGLSSQSDILTTQQRDTMQHNLKIQOEMAELEAVLHQHQSQPSNSYPSIISDSSALE 1440
QY 1441 DLRNPEQSTSEKAVLTSQKSSEYPIQONPEGLSADKFEVSADSSTSKNKEPGVERSSPSK 1500
Db 1441 DLRNPEQSTSEKAVLTSQKSSEYPIQONPEGLSADKFEVSADSSTSKNKEPGVERSSPSK 1500
QY 1501 CPSLDDRWTFMHSQSGSLQNRNYPQBELTKVVDVVEQOOLEESGPHDLTSTYLPQDLEG 1560
Db 1501 CPSLDDRWTFMHSQSGSLQNRNYPQBELTKVVDVVEQOOLEESGPHDLTSTYLPQDLEG 1560
QY 1561 TPYLESGISLFDSDPDSPEDRAPESARVGNIPSTSAKYPQLKVAESAQSPAAAHPT 1620
Db 1561 TPYLESGISLFDSDPDSPEDRAPESARVGNIPSTSAKYPQLKVAESAQSPAAAHPT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEEPMVYKFAKHHITLTNLI 1680
Db 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEEPMVYKFAKHHITLTNLI 1680
QY 1681 TETTHVMKTDAEFCERTLAKYELGIAGGKVVSVFWTQSTIKERKMLNEHDFEVRGDV 1740
Db 1681 TETTHVMKTDAEFCERTLAKYELGIAGGKVVSVFWTQSTIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHQPKRAESODRKIFRGLEICCCGPTNPTDLEWVVLGCASVVKELSSFTL 1800
Db 1741 VNGRNHQPKRAESODRKIFRGLEICCCGPTNPTDLEWVVLGCASVVKELSSFTL 1800
QY 1801 GTGVHPVVVQPDWNTDNGFHAIGQMCBAPVVTREWLDLSVALYQCQELDTYLIPQIPH 1860

Db 1801 GTGVHPVVVQPDWNTDNGFHAIGQMCBAPVVTREWLDLSVALYQCQELDTYLIPQIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863
RESULT 10
PCT-US95-10202-2
; Sequence 2, Application PC/TUS9510202
; GENERAL INFORMATION:
; APPLICANT: Shattuck-Eidens, Donna M.
; APPLICANT: Simard, Jacques
; APPLICANT: Emi, Mitsuru
; APPLICANT: Nakamura, Yusuke
; APPLICANT: Durocher, Francine
; TITLE OF INVENTION: In Vivo Mutations and Polymorphisms
; TITLE OF INVENTION: in the 17q-Linked Breast and Ovarian Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
; STREET: 1201 New York Avenue, N.W., Suite 1000
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10202
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,305
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,824
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08-308,104
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/300,266
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,221
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-109347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-10202-2
Query Match 99.9%; Score 9642; DB 5; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MDLSALRVEEVQVNIAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNKKGPSQ 60
Db 1 MDLSALRVEEVQVNIAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNKKGPSQ 60
QY 61 CPLCKNDITKKSLOBSTFSQLVBEELKIIICAFOLDTGLEYANSYNFAKKENNSPEHLKD 120
Db 61 CPLCKNDITKKSLOBSTFSQLVBEELKIIICAFOLDTGLEYANSYNFAKKENNSPEHLKD 120
QY 121 EYSIIQSOGYRNRAKRLQSPENPSLQETSLSVQLSNLGVTRLTQRKIQPOKTSVYI 180
Db 121 EYSIIQSOGYRNRAKRLQSPENPSLQETSLSVQLSNLGVTRLTQRKIQPOKTSVYI 180
QY 181 ELGSDSSEDVTNKATYCSVGQDELQIITPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDVTNKATYCSVGQDELQIITPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
QY 241 PSNNDLNTTERAAARRHPKTYQGSVSMLHVEPCGTNTHASSLOHENSLLLTKDRMNVE 300
Db 241 PSNNDLNTTERAAARRHPKTYQGSVSMLHVEPCGTNTHASSLOHENSLLLTKDRMNVE 300
QY 301 KAEFCNKSQKPLARSOHNRWAGSKETCNDRRTPSTEKKVDLNLADPLCERKEWNKQKLP 360
Db 301 KAEFCNKSQKPLARSOHNRWAGSKETCNDRRTPSTEKKVDLNLADPLCERKEWNKQKLP 360
QY 361 SENPRDTEVPWITLNSIQVNEWFSRSDLLGSDSDHGESESNAKVADVLDVLNEVD 420
Db 361 SENPRDTEVPWITLNSIQVNEWFSRSDLLGSDSDHGESESNAKVADVLDVLNEVD 420
QY 421 EYSGSSEKIDLLADPHALCKSERVHKSVSNEIEDKIFGKYRKKASLPNLSHVTEN 480
Db 421 EYSGSSEKIDLLADPHALCKSERVHKSVSNEIEDKIFGKYRKKASLPNLSHVTEN 480
QY 481 LIIGAFVTEPOLIOERPLTNLKKRPTSLGHPEDFIKKADLAVOKTPEMINOGTNOTE 540
Db 481 LIIGAFVTEPOLIOERPLTNLKKRPTSLGHPEDFIKKADLAVOKTPEMINOGTNOTE 540
QY 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNMELENI 600
Db 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNMELENI 600
QY 601 HNSKAPKNRLRRKSSRTHIALVELVSRNLSPNCNCTELQIDSCSSSEEEIKKKYNOMP 660
Db 601 HNSKAPKNRLRRKSSRTHIALVELVSRNLSPNCNCTELQIDSCSSSEEEIKKKYNOMP 660
QY 661 RHSRLQLMECKEPATGAKKNKNEOTSKRHSDTPELKLTNAPGSFTKCSNTSELKE 720
Db 661 RHSRLQLMECKEPATGAKKNKNEOTSKRHSDTPELKLTNAPGSFTKCSNTSELKE 720
QY 721 FVNPSLPREEKEEKLTVKVSNNNAEDPKDMLSGERVLOTERSVESSISILVPGTDYGTQ 780
Db 721 FVNPSLPREEKEEKLTVKVSNNNAEDPKDMLSGERVLOTERSVESSISILVPGTDYGTQ 780
QY 781 ESIISLLEVSTLIGKAKTEPNKCVSOCAAFENPKGLIHGCSKDNDRNDETFGPKYPLGHEVNH 840
Db 781 ESIISLLEVSTLIGKAKTEPNKCVSOCAAFENPKGLIHGCSKDNDRNDETFGPKYPLGHEVNH 840
QY 841 RETSIEMESELDAQYLQNTFKVSKROSFAFVSNPGNAEEECATFSAHSGSLKKOSPKVT 900
Db 841 RETSIEMESELDAQYLQNTFKVSKROSFAFVSNPGNAEEECATFSAHSGSLKKOSPKVT 900
QY 901 FECEQKEENQCKNESNTKPVQTVNITAGFPVVGQKDKPVDNAKCSIKGGSRFCLSSQFRG 960
Db 901 FECEQKEENQCKNESNTKPVQTVNITAGFPVVGQKDKPVDNAKCSIKGGSRFCLSSQFRG 960
QY 961 NETGLITPNKKGILLQNPYRPPFPKSFVKTKCKKNLLEENFEHSMSPEREMGNENIP 1020
Db 961 NETGLITPNKKGILLQNPYRPPFPKSFVKTKCKKNLLEENFEHSMSPEREMGNENIP 1020
QY 1021 STVSTISRNNTRENVPFEASSNNNEVGSSTNEVSGSSINEIGSSDENIQAEIQRNRPKL 1080
Db 1021 STVSTISRNNTRENVPFEASSNNNEVGSSTNEVSGSSINEIGSSDENIQAEIQRNRPKL 1080
QY 1081 NAMLRGLVQPEVYKQSLPGSNCKHPEIKKQYEVVQTVNTDFSPYLLISDNLEQPMGSS 1140

Db 1081 NAMLRGLVQPEVYKQSLPGSNCKHPEIKKQYEVVQTVNTDFSPYLLISDNLEQPMGSS 1140
QY 1141 HASQVCSETPDDLLDGDGEIKEIDTFSFAENDIKESSAVSKSVOKGELSRSPPFTHTHLAQ 1200
Db 1141 HASQVCSETPDDLLDGDGEIKEIDTFSFAENDIKESSAVSKSVOKGELSRSPPFTHTHLAQ 1200
QY 1201 GYRRGAKKLESSEENLSSEDEELPCFQHLHFGKVNNTPSQSTRHSTVATECLSKNTEENL 1260
Db 1201 GYRRGAKKLESSEENLSSEDEELPCFQHLHFGKVNNTPSQSTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDGCSNOVILAKASQEHHLSEETKCSASFSSQCSSELEDLTANTNTQDPFLIGS 1320
Db 1261 LSLKNSLNDGCSNOVILAKASQEHHLSEETKCSASFSSQCSSELEDLTANTNTQDPFLIGS 1320
QY 1321 SKQMRHOSQSGVGLSDKELVSDDEERGTGLEFENNQBQESMDNLGEAASCESETSYSE 1380
Db 1321 SKQMRHOSQSGVGLSDKELVSDDEERGTGLEFENNQBQESMDNLGEAASCESETSYSE 1380
QY 1381 DCSGLSSQSDILITTOQRDTMHOHNLIKLQOEMAELEAVLEQHGQSPSNSYPISIISSSSALE 1440
Db 1381 DCSGLSSQSDILITTOQRDTMHOHNLIKLQOEMAELEAVLEQHGQSPSNSYPISIISSSSALE 1440
QY 1441 DLRNPEQSTSEKAVLTQSSEYPISONPEGLSADKFEVSADSSSTSKNKEPVERSSPSK 1500
Db 1441 DLRNPEQSTSEKAVLTQSSEYPISONPEGLSADKFEVSADSSSTSKNKEPVERSSPSK 1500
QY 1501 CPSLDDRWYMHSCSGSLQNRNYSQEBELIKVVDVEBQOLESBPHDLTETSYLPRQDLEG 1560
Db 1501 CPSLDDRWYMHSCSGSLQNRNYSQEBELIKVVDVEBQOLESBPHDLTETSYLPRQDLEG 1560
QY 1561 TPYLESIGSILFSDDPESDPESTRAPESARVGNIPSSTSALKVPOLKVAESAQSPAAAHTT 1620
Db 1561 TPYLESIGSILFSDDPESDPESTRAPESARVGNIPSSTSALKVPOLKVAESAQSPAAAHTT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEEFMLVYKPAKHHHTLTNLI 1680
Db 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEEFMLVYKPAKHHHTLTNLI 1680
QY 1681 TEETHVVMKTDAEFVCERTLKYPFLGTTAGKVVVSYFWVTQSIKERKMLNEHDFEVRGDV 1740
Db 1681 TEETHVVMKTDAEFVCERTLKYPFLGTTAGKVVVSYFWVTQSIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHQGPARRAESQDRKIFRGLIEICCYGPFNTMPTDQLEWMVQLCGASVVKELSSFTL 1800
Db 1741 VNGRNHQGPARRAESQDRKIFRGLIEICCYGPFNTMPTDQLEWMVQLCGASVVKELSSFTL 1800
QY 1801 GTGVHPITVVVQPDWNTEDNGFHAIGQMCAPVVTREWVLDVSVALYQCOELDTYLIQIPH 1860
Db 1801 GTGVHPITVVVQPDWNTEDNGFHAIGQMCAPVVTREWVLDVSVALYQCOELDTYLIQIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863
RESULT 11
PCT-US95-10203-2
; Sequence 2, Application PC/TUS9510203
; GENERAL INFORMATION:
; APPLICANT: Skolnick, Mark H.
; APPLICANT: Goldgar, David E.
; APPLICANT: Miki, Yoshio
; APPLICANT: Swenson, Jeff
; APPLICANT: Kamb, Alexander
; APPLICANT: Harshman, Keith D.
; APPLICANT: Shattuck-Bidens, Donna M.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Wiseman, Roger W.
; APPLICANT: Futreal, P. Andrew
; TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
; TITLE OF INVENTION: Susceptibility Gene
; NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
 STREET: 1201 New York Avenue, N.W., Suite 1000
 CITY: Washington
 STATE: DC
 COUNTRY: USA
 ZIP: 20005

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/10203
 FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US

FILING DATE: 07-JUN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/409,305

FILING DATE: 24-MAR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/348,824

FILING DATE: 29-NOV-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08-308,104

FILING DATE: 16-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/300,266

FILING DATE: 02-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/289,221

FILING DATE: 12-AUG-1994

ATTORNEY/AGENT INFORMATION:

NAME: Innen, Jeffrey L.

REGISTRATION NUMBER: 28,957

REFERENCE/DOCKET NUMBER: 24884-109347

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-962-4810

TELEFAX: 202-962-8300

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1863 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US95-10203-2

Query Match 99.9%; Score 9642; DB 5; Length 1863;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	MDLSALRVEEVQNVINAMQKILECPICLLEIKPEVSTKCDHIFCKFCMLKLLNQKKGPSQ	60
Db	1	MDLSALRVEEVQNVINAMQKILECPICLLEIKPEVSTKCDHIFCKFCMLKLLNQKKGPSQ	60
QY	61	CPICKNDITKRSLOESTRFSQVVEELKLIICAFQDITGLYANSYNFAKKNENSPHEHLD	120
Db	61	CPICKNDITKRSLOESTRFSQVVEELKLIICAFQDITGLYANSYNFAKKNENSPHEHLD	120
QY	121	EVSIIQSMGYRNRAKRLQSEPNPSLQETSLSVQLSNLGTVRTLRKRIOPQKTSVYI	180
Db	121	EVSIIQSMGYRNRAKRLQSEPNPSLQETSLSVQLSNLGTVRTLRKRIOPQKTSVYI	180
QY	181	ELGSDSDSDTVNKATYCSVDQELLQITPQGTDRDEISLDSAKKAACEFSETDVTNTEHHQ	240
Db	181	ELGSDSDSDTVNKATYCSVDQELLQITPQGTDRDEISLDSAKKAACEFSETDVTNTEHHQ	240
QY	241	PSNNDLNTTEKRAAERHPEKYOGSSVSNLHVPCGNTNTHASSLQHNENSLLLTKDRMNYE	300
Db	241	PSNNDLNTTEKRAAERHPEKYOGSSVSNLHVPCGNTNTHASSLQHNENSLLLTKDRMNYE	300

QY	301	KAFCNCKSQPGLARSOHNRWAGSKETCNDRRPTSTPEKKVVDLNADPLCERKWNKQKLP	360
Db	301	KAFCNCKSQPGLARSOHNRWAGSKETCNDRRPTSTPEKKVVDLNADPLCERKWNKQKLP	360
QY	361	SENPRDTEVPWITLNSSIQKVNEWFSRDELLGSDSDSHDGESENAKVADVLDVLYNEVD	420
Db	361	SENPRDTEVPWITLNSSIQKVNEWFSRDELLGSDSDSHDGESENAKVADVLDVLYNEVD	420
QY	421	EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRKKASLPNLSHVTEN	480
Db	421	EYSGSSEKIDLLASDPHEALICKSERVHKSVESNIEDKIFGKTYRKKASLPNLSHVTEN	480
QY	481	LIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPDEFIKKADLAVQKTPEMINQGTNOTE	540
Db	481	LIIGAFVTEPQIIQERPLTNKLRKRPTSGLHPDEFIKKADLAVQKTPEMINQGTNOTE	540
QY	541	QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINMELELNI	600
Db	541	QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISISSINMELELNI	600
QY	601	HNSKAPKKNRLRRKSSSTRHIALELVVSRLNLPNPPNCTELQIDSCSSSEETKKKYNQMPV	660
Db	601	HNSKAPKKNRLRRKSSSTRHIALELVVSRLNLPNPPNCTELQIDSCSSSEETKKKYNQMPV	660
QY	661	RHSRLQLMEGKEPATGAKSKNKPNEQTSKRHDSDTFFPELKLNTNAPGSFTKCSNTSELKE	720
Db	661	RHSRLQLMEGKEPATGAKSKNKPNEQTSKRHDSDTFFPELKLNTNAPGSFTKCSNTSELKE	720
QY	721	FVNPSLPREEKEKLETVKVSNNADDPKDLMLSGERVLOTERSVSSSSISLVPGTDYGTQ	780
Db	721	FVNPSLPREEKEKLETVKVSNNADDPKDLMLSGERVLOTERSVSSSSISLVPGTDYGTQ	780
QY	781	ESISLLEVTGLKAKTEPNKCVSOCAAFENPKGLIHGCCSKDNNDTEGKYPLGHEVNIHS	840
Db	781	ESISLLEVTGLKAKTEPNKCVSOCAAFENPKGLIHGCCSKDNNDTEGKYPLGHEVNIHS	840
QY	841	RETSIEMESELDAQYLQNTFKVSKRQSFALFSPNPGNAEECATFSAHSGSLKKQSPKVT	900
Db	841	RETSIEMESELDAQYLQNTFKVSKRQSFALFSPNPGNAEECATFSAHSGSLKKQSPKVT	900
QY	901	FECEQKEENQGNESNIKPVQTVNITAGPPVVGQDKDPVDNAKCSIKGSRCLSSQFRG	960
Db	901	FECEQKEENQGNESNIKPVQTVNITAGPPVVGQDKDPVDNAKCSIKGSRCLSSQFRG	960
QY	961	NETGLITPNKHGLLQNPYRIPPLFPIKSVKTKCKKNLLLEENFEHSMSPEREMNGENIP	1020
Db	961	NETGLITPNKHGLLQNPYRIPPLFPIKSVKTKCKKNLLLEENFEHSMSPEREMNGENIP	1020
QY	1021	STVSTISRNIIRNVFKEASSNINEVGSSTNEVGSSSINEIGSSDENIOAELGRNRPKL	1080
Db	1021	STVSTISRNIIRNVFKEASSNINEVGSSTNEVGSSSINEIGSSDENIOAELGRNRPKL	1080
QY	1081	NAMLRLGVLOPEVYKOSLPGSNCKHPEIKKQYEEVVQVNTDFSPYLISDNLEQPMGSS	1140
Db	1081	NAMLRLGVLOPEVYKOSLPGSNCKHPEIKKQYEEVVQVNTDFSPYLISDNLEQPMGSS	1140
QY	1141	HASQVCSETPDLLDDGEIKEDTSAENDIKESSAVFSKVQKGLSRSPSFTHHLAQ	1200
Db	1141	HASQVCSETPDLLDDGEIKEDTSAENDIKESSAVFSKVQKGLSRSPSFTHHLAQ	1200
QY	1201	GYRRGAKKLESSEENLSSEDEELPCFQHLFGKVNNIPSQSTRHSTVATECLSKNTEENL	1260
Db	1201	GYRRGAKKLESSEENLSSEDEELPCFQHLFGKVNNIPSQSTRHSTVATECLSKNTEENL	1260
QY	1261	LSLKNSLNDCSNQVILAKASQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS	1320
Db	1261	LSLKNSLNDCSNQVILAKASQEHHLSEETKCSASLFSQCSSELEDLTANTNTQDPFLIGS	1320
QY	1321	SKQMRHQSESQGVGLSDKELVSDDEERGTCLEENQEQESMDNSNLGEAASGESETSVSE	1380
Db	1321	SKQMRHQSESQGVGLSDKELVSDDEERGTCLEENQEQESMDNSNLGEAASGESETSVSE	1380
QY	1381	DCSGLSSQSDILTTQOQRTMQHNLIKLOQEMAELEAVLBOHGSQPSNSYPSIISDSSALE	1440

```

1381 DCSGLSSQSDILLTQORDTMOHNLKLOQMAEAVLEQHQSPNSYPSIISDSSALE 1440
1441 DURNPQSTSEKAVLTSQKSSYPIISQNEGLSADKFEVSADSTSKNKEGVERSSPSK 1500
1441 DURNPQSTSEKAVLTSQKSSYPIISQNEGLSADKFEVSADSTSKNKEGVERSSPSK 1500
1501 CFSLDNRWYMHSCGSLQNRNYPQOEELIKVVDVEQOEESGPHDLTETSYLPRQDLEG 1560
1501 CFSLDNRWYMHSCGSLQNRNYPQOEELIKVVDVEQOEESGPHDLTETSYLPRQDLEG 1560
1561 TPYLESIGISLSDPESDSEDAPESARVGNIPSTGALKVPOLKVAESAQSPAAAHHT 1620
1561 TPYLESIGISLSDPESDSEDAPESARVGNIPSTGALKVPOLKVAESAQSPAAAHHT 1620
1621 DTAGYNAMESVSRKPELTASTERVNKRMSVMVSGSLTPEEFMLVYKFAKHHITLNL 1680
1621 DTAGYNAMESVSRKPELTASTERVNKRMSVMVSGSLTPEEFMLVYKFAKHHITLNL 1680
1681 TEETHVMKTDAEFCERTLYFLGIAGGKVVVSWYFWVTQSIKERKMLNEDHFEVRGDV 1740
1681 TEETHVMKTDAEFCERTLYFLGIAGGKVVVSWYFWVTQSIKERKMLNEDHFEVRGDV 1740
1741 VNGRNHQPRAESQDRKIFRGLICCGYPTNNMPTDQLEMMVQLCGASVVKELSSFTL 1800
1741 VNGRNHQPRAESQDRKIFRGLICCGYPTNNMPTDQLEMMVQLCGASVVKELSSFTL 1800
1801 GTGVHPVIVVQDANTEONGFHAIGOMCEAPVVTREWVLDSVALYQOEELTYLIPQIPH 1860
1801 GTGVHPVIVVQDANTEONGFHAIGOMCEAPVVTREWVLDSVALYQOEELTYLIPQIPH 1860
1861 SHY 1863
1861 SHY 1863

```

RESULT 12

```

PCT-US95-10220-2
Sequence 2. Application PC/TUS9510220
GENERAL INFORMATION:
APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavtigian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: Method for diagnosing a
TITLE OF INVENTION: Predisposition for Breast and Ovarian Cancer
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESS: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10220
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:

```

```

APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08-308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-10220-2

```

```

Query Match 99.9%; Score 9642; DB 5; Length 1863;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1862; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 MDLSALRVEEVQNVINAMQKILECPICILELKEPVSTKCDHIFCKFCKMLNKKGPSQ 60
DB 1 MDLSALRVEEVQNVINAMQKILECPICILELKEPVSTKCDHIFCKFCKMLNKKGPSQ 60
QY 61 CPICKNDITKRSLOESTRFESOLVVEELKIIICAFOLDTCLEVANSYNFAKKENNSPEHLKD 120
DB 61 CPICKNDITKRSLOESTRFESOLVVEELKIIICAFOLDTCLEVANSYNFAKKENNSPEHLKD 120
QY 121 EVSIIQSMGYNRNRAKRLQSEPNPSLQETSLSVQLSNLGTVRTLTQRIQPKQTSVYI 180
DB 121 EVSIIQSMGYNRNRAKRLQSEPNPSLQETSLSVQLSNLGTVRTLTQRIQPKQTSVYI 180
QY 181 ELGSDSSSDIVNKATYCSVGDOELLQITPQGTREISLDSAKKAACERSETDVTNTEHHQ 240
DB 181 ELGSDSSSDIVNKATYCSVGDOELLQITPQGTREISLDSAKKAACERSETDVTNTEHHQ 240
QY 241 PSNNDLNTTEKRAAERHPEKYQGVSSVNLHVEPCGTNTTHASSLOHENSLLLTDRMNV 300
DB 241 PSNNDLNTTEKRAAERHPEKYQGVSSVNLHVEPCGTNTTHASSLOHENSLLLTDRMNV 300
QY 301 KAFCNKSKQFGLARSQHRNWSKTCNDRRTPTSTKVKVDNADPLCERKEWNKKQLPC 360
DB 301 KAFCNKSKQFGLARSQHRNWSKTCNDRRTPTSTKVKVDNADPLCERKEWNKKQLPC 360
QY 361 SENPRDTEVPWITLNSSIQKVNEWFSRDELGSDSDSHGESNNAKVADVLVDNEVD 420
DB 361 SENPRDTEVPWITLNSSIQKVNEWFSRDELGSDSDSHGESNNAKVADVLVDNEVD 420
QY 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKKASLPNLSHVTE 480
DB 421 EYSGSSEKIDLLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKKASLPNLSHVTE 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPMINOQTNQTE 540
DB 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPMINOQTNQTE 540
QY 541 QNGOVNMITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPITSSSISNMELELNI 600
DB 541 QNGOVNMITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPITSSSISNMELELNI 600

```

QY 601 HNSKAPKNNLRKSSRRHHHALELVVSRNLSPNCTELQIDSCSSSEEEKKKYNQMPV 660
Db 601 HNSKAPKNNLRKSSRRHHHALELVVSRNLSPNCTELQIDSCSSSEEEKKKYNQMPV 660
QY 661 RHRNQLMEGKKEPATCAKSNKPNQTSKRHSDTFPELKLTPNAGSFTKCSNTSELKE 720
Db 661 RHRNQLMEGKKEPATCAKSNKPNQTSKRHSDTFPELKLTPNAGSFTKCSNTSELKE 720
QY 721 FVNPSPREEKEKLETVKVSNNAEADPKDMLSGERVLOQTERSVESISSLVPGTDYGTQ 780
Db 721 FVNPSPREEKEKLETVKVSNNAEADPKDMLSGERVLOQTERSVESISSLVPGTDYGTQ 780
QY 781 ESISLLEVSTLGAKTEPNKVCQCAAFENPKGLIHGCKDNNDTEGFKYPLGHEVNHHS 840
Db 781 ESISLLEVSTLGAKTEPNKVCQCAAFENPKGLIHGCKDNNDTEGFKYPLGHEVNHHS 840
QY 841 RETSIEMEESELDAQYLQNTFKVSKQSFALFSPNAGNAEEECATFSAHSGSLKKQSPKVT 900
Db 841 RETSIEMEESELDAQYLQNTFKVSKQSFALFSPNAGNAEEECATFSAHSGSLKKQSPKVT 900
QY 901 FECEQKEENGKGNESNIKPVQTVNITAGFPVVGQKQKPDVNAKCSIKGSRFCLSSQFRG 960
Db 901 FECEQKEENGKGNESNIKPVQTVNITAGFPVVGQKQKPDVNAKCSIKGSRFCLSSQFRG 960
QY 961 NETGLITPNKHGLLQNPYRIPPLFPFKTSFVKTKCKKNLLEENFEHSMSPEREMGNENIP 1020
Db 961 NETGLITPNKHGLLQNPYRIPPLFPFKTSFVKTKCKKNLLEENFEHSMSPEREMGNENIP 1020
QY 1021 STVSTISRNNIRENVKFASSSNI NEVGSTNEVGSSINIGSSDENIQAEELNRGPKL 1080
Db 1021 STVSTISRNNIRENVKFASSSNI NEVGSTNEVGSSINIGSSDENIQAEELNRGPKL 1080
QY 1081 NAMLRGLVLQPEYVKOSLPGSNCKHPEIKOEYEEVQTVNTDFSPVYLSNDLEQPMGSS 1140
Db 1081 NAMLRGLVLQPEYVKOSLPGSNCKHPEIKOEYEEVQTVNTDFSPVYLSNDLEQPMGSS 1140
QY 1141 HASQVCSETPDDLLDGEIKEDTSAFENDIKESSAVFSKVQKGLSRSPFTHTHQAQ 1200
Db 1141 HASQVCSETPDDLLDGEIKEDTSAFENDIKESSAVFSKVQKGLSRSPFTHTHQAQ 1200
QY 1201 GYRRGAKKLESSEENLSDEDELPCQHLHFGKVNIPQSOTRHSIVATECLSKNTEENL 1260
Db 1201 GYRRGAKKLESSEENLSDEDELPCQHLHFGKVNIPQSOTRHSIVATECLSKNTEENL 1260
QY 1261 LSLKNSLDCSNQVILAKASQEHHLSEETKCSASLFSQCSSELEDLTANTNTODPFLIGS 1320
Db 1261 LSLKNSLDCSNQVILAKASQEHHLSEETKCSASLFSQCSSELEDLTANTNTODPFLIGS 1320
QY 1321 SKQMRHQSESQGVGLSDKELVSDDEERGTLGLENNOEQSMDSNLGEAASGCESETSVSE 1380
Db 1321 SKQMRHQSESQGVGLSDKELVSDDEERGTLGLENNOEQSMDSNLGEAASGCESETSVSE 1380
QY 1381 DCSGLSSQSDILTQORDTMOHNLIKLOQEMAELEAVLEOHGQSPNSYPSIIISDSALE 1440
Db 1381 DCSGLSSQSDILTQORDTMOHNLIKLOQEMAELEAVLEOHGQSPNSYPSIIISDSALE 1440
QY 1441 DLNRPQSTSEKAVLTSQKSSEYPIQSONPEGLSADKFEVSADSTSKNKEPVGVERSPSK 1500
Db 1441 DLNRPQSTSEKAVLTSQKSSEYPIQSONPEGLSADKFEVSADSTSKNKEPVGVERSPSK 1500
QY 1501 CPSLDRWYMHSCGSLQNRNYPQBELIKVVDVEEQLEESGPHDLTETSYLPRQDLEG 1560
Db 1501 CPSLDRWYMHSCGSLQNRNYPQBELIKVVDVEEQLEESGPHDLTETSYLPRQDLEG 1560
QY 1561 TPYLESGLISLFDPSDDPSDPRAPESARVGNIPSTSAKVPQLKVAESAQAHAHTT 1620
Db 1561 TPYLESGLISLFDPSDDPSDPRAPESARVGNIPSTSAKVPQLKVAESAQAHAHTT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERYNKRMSMVVSGLTPEERMLVYKFAKKHHTITNL 1680
Db 1621 DTAGYNAMESVSREKPELTASTERYNKRMSMVVSGLTPEERMLVYKFAKKHHTITNL 1680
QY 1681 TEETHVWVKTDAEFVCERTLKYFLGIAGGKWWVSYFWVTQSIKERKMLNEHDFEVRGDV 1740

Db 1681 TEETHVWVKTDAEFVCERTLKYFLGIAGGKWWVSYFWVTQSIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHOGPKRARESODRKIFRGLIEICCCYGPFTNMPTDQLEWNVOLCGASVVKELSSFTL 1800
Db 1741 VNGRNHOGPKRARESODRKIFRGLIEICCCYGPFTNMPTDQLEWNVOLCGASVVKELSSFTL 1800
QY 1801 GTGVHPITVVVQPDWATEDNGFHAIGOMCEAPVVTREWLDSVALYOCQELDTYLIQPIPH 1860
Db 1801 GTGVHPITVVVQPDWATEDNGFHAIGOMCEAPVVTREWLDSVALYOCQELDTYLIQPIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863

RESULT 13
US-08-598-591-2
; Sequence 2, Application US/08598591
; Patent No. 5654155
; GENERAL INFORMATION:
; APPLICANT: Allen, Antonette C.
; APPLICANT: Alvares, Christopher P.
; APPLICANT: Critz, Brenda S.
; APPLICANT: Murphy Patricia D.
; APPLICANT: Olson, Sheri J.
; APPLICANT: Schelter, Denise B.
; APPLICANT: Zeng, Bin
; TITLE OF INVENTION: A Consensus Sequence of the Human BRCA1 Gene
; Patent No. 5654155
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: 699 Prince St.
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/598,591
; FILING DATE: herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Swecker, Robert S.
; REGISTRATION NUMBER: 19,885
; REFERENCE/DOCKET NUMBER: 020160-282
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCA1
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17
; MAP POSITION: 17q21
US-08-598-591-2

Query Match 99.9%; Score 9635; DB 1; Length 1863;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1860; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MDLSALRVEEVQVNIAMQKILCEPICLELBIKEPVSTKCDHIFCKFCMKLLNKKGPSQ 60
Db 1 MDLSALRVEEVQVNIAMQKILCEPICLELBIKEPVSTKCDHIFCKFCMKLLNKKGPSQ 60
QY 61 CPLCKNDITKRSLOESTRESOLVEELLKIICAFQOLDTGLEYANSYNFAKKNNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRESOLVEELLKIICAFQOLDTGLEYANSYNFAKKNNSPEHLKD 120
QY 121 EVSIIQSMGYRNRKRLQSEPNPSLOETSLVSOLSNLGTVRTLTKORTQOPKTSYVI 180
Db 121 EVSIIQSMGYRNRKRLQSEPNPSLOETSLVSOLSNLGTVRTLTKORTQOPKTSYVI 180
QY 181 ELGSDSEDTVNAKYCSVGQDQLLQITPQGTREDEISLDSAKKACESEFSETDVTNTEHHQ 240
Db 181 ELGSDSEDTVNAKYCSVGQDQLLQITPQGTREDEISLDSAKKACESEFSETDVTNTEHHQ 240
QY 241 PSNNDLTTEKRAAERHPKYOSSVSNLHVPCGNTNTHASSLOHENSLLLTQDRMVE 300
Db 241 PSNNDLTTEKRAAERHPKYOSSVSNLHVPCGNTNTHASSLOHENSLLLTQDRMVE 300
QY 301 KAEFCNKSQOPGLARSOHNRWAGSETCNDRTPTSTKVKVDLNADPLCERKWNKQKLPC 360
Db 301 KAEFCNKSQOPGLARSOHNRWAGSETCNDRTPTSTKVKVDLNADPLCERKWNKQKLPC 360
QY 361 SENPRDTPWITLNSSIQKYNFWSRDELGLGSDSDHGESESNAKVADVLDVINEVD 420
Db 361 SENPRDTPWITLNSSIQKYNFWSRDELGLGSDSDHGESESNAKVADVLDVINEVD 420
QY 421 EYSGSEKIDLLASDPHEALICKSERVHSKVESNIEDKIFGKTYRKKASIPNLSHVTE 480
Db 421 EYSGSEKIDLLASDPHEALICKSERVHSKVESNIEDKIFGKTYRKKASIPNLSHVTE 480
QY 481 LIIGAEVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVOKTPEMINOCTNOTE 540
Db 481 LIIGAEVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVOKTPEMINOCTNOTE 540
QY 541 QNGQVNNITNSGHENKTGDSIQNEKNPNP IESLEKESAFKTKABPISSISNMELELNI 600
Db 541 QNGQVNNITNSGHENKTGDSIQNEKNPNP IESLEKESAFKTKABPISSISNMELELNI 600
QY 601 HNSKAPKKRLRKSTRIIHALELVSRNLSPNCTELQIDSCSSSEIEIKKKYNQMPV 660
Db 601 HNSKAPKKRLRKSTRIIHALELVSRNLSPNCTELQIDSCSSSEIEIKKKYNQMPV 660
QY 661 RISRNLQMEGKEPATGAKSKNPNQETSKRHDSDFTELKLTNAPGSFTKCSNTSELKE 720
Db 661 RISRNLQMEGKEPATGAKSKNPNQETSKRHDSDFTELKLTNAPGSFTKCSNTSELKE 720
QY 721 FVNPSLPREEKEETLVKVSNNADPKDLMLSGERVLOTERSVESSISLVPCTDYGTQ 780
Db 721 FVNPSLPREEKEETLVKVSNNADPKDLMLSGERVLOTERSVESSISLVPCTDYGTQ 780
QY 781 ESISLLEVTSLGKATEPNKCVSQCAAFENPKGLIHGCKDNRNDTEGFKYPLGHEVNH 840
Db 781 ESISLLEVTSLGKATEPNKCVSQCAAFENPKGLIHGCKDNRNDTEGFKYPLGHEVNH 840
QY 841 RETSIEMESELDAQYLONTFKVSRQSFALFSPNGNAEECATFSAHSGSLKKOSPKVT 900
Db 841 RETSIEMESELDAQYLONTFKVSRQSFALFSPNGNAEECATFSAHSGSLKKOSPKVT 900
QY 901 FCEQKEENQGNESNIPKPVQTNITAGFPVVGQKDPVDNAKCSIKGSRFCLSSQFRG 960
Db 901 FCEQKEENQGNESNIPKPVQTNITAGFPVVGQKDPVDNAKCSIKGSRFCLSSQFRG 960
QY 961 NETGLITPNKHGLLQNPRIPLPFKISFVTKCKKNLLEENFEHSHNSPSPRENGENIP 1020
Db 961 NETGLITPNKHGLLQNPRIPLPFKISFVTKCKKNLLEENFEHSHNSPSPRENGENIP 1020
QY 1021 STVSTISRNNIRENVFKKASSNINEVSSSTNEVSSSTNEIGSSDENIQABLGRNRPKL 1080
Db 1021 STVSTISRNNIRENVFKKASSNINEVSSSTNEVSSSTNEIGSSDENIQABLGRNRPKL 1080
QY 1081 NAMLRGLVQPEVYKQSLPGSNCKHPEIKKOEYEVVQTVNTDFSPYLISDNLBPQMGSS 1140

Db 1081 NAMLRGLVQPEVYKQSLPGSNCKHPEIKKOEYEVVQTVNTDFSPYLISDNLBPQMGSS 1140
QY 1141 HASQVCSETPDLLDDGGEIKEDTSPAENDIKESSAVFSKSVQKGLSRSPSFTHTHAQ 1200
Db 1141 HASQVCSETPDLLDDGGEIKEDTSPAENDIKESSAVFSKSVQKGLSRSPSFTHTHAQ 1200
QY 1201 GYRRGAKKLESSEENLSEDEELPCFQHLFGKVNINISQSTRHSTVATECLSKNTEENL 1260
Db 1201 GYRRGAKKLESSEENLSEDEELPCFQHLFGKVNINISQSTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDSCNQVILAKASQEHHLSEETKCSASFSSQCSSELEDLTANTNTQDPFLIGS 1320
Db 1261 LSLKNSLNDSCNQVILAKASQEHHLSEETKCSASFSSQCSSELEDLTANTNTQDPFLIGS 1320
QY 1321 SKQMRHOSQSGVGLSDKELVSDDEERTGLLENNQEQSDMSNLGEAASGESETSVSE 1380
Db 1321 SKQMRHOSQSGVGLSDKELVSDDEERTGLLENNQEQSDMSNLGEAASGESETSVSE 1380
QY 1381 DCSGLSSQSDILTQOORDTMOHNLIKLOEAMAELEAVLEOHGSOQPSNSYPSIISDSSALE 1440
Db 1381 DCSGLSSQSDILTQOORDTMOHNLIKLOEAMAELEAVLEOHGSOQPSNSYPSIISDSSALE 1440
QY 1441 DLNPEQSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSSSTKNKEPGEVERSSPSK 1500
Db 1441 DLNPEQSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSSSTKNKEPGEVERSSPSK 1500
QY 1501 CPSLDDRRWMHSCSGSLQNRNYPQOEELIKVVDVEEQOLESQPHDLTETSYLPRQDLEG 1560
Db 1501 CPSLDDRRWMHSCSGSLQNRNYPQOEELIKVVDVEEQOLESQPHDLTETSYLPRQDLEG 1560
QY 1561 TPYLESIGLSFSDDPSPEDRAPESARVGNIPSSTSALKVPOLKVAESAQSPAAAHTT 1620
Db 1561 TPYLESIGLSFSDDPSPEDRAPESARVGNIPSSTSALKVPOLKVAESAQSPAAAHTT 1620
QY 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEBFMLYKFKARKHHTLTNLI 1680
Db 1621 DTAGYNAMESVSREKPELTASTERVNKRMSVMVSGLTPEBFMLYKFKARKHHTLTNLI 1680
QY 1681 TEETHVVMKTDAEFVCERTLYFLGTAGGKVVYFVWVTOSIKERKMLNEHDFEVRGDV 1740
Db 1681 TEETHVVMKTDAEFVCERTLYFLGTAGGKVVYFVWVTOSIKERKMLNEHDFEVRGDV 1740
QY 1741 VNGRNHQPKRARSQDRKIFRGLIEICCYGPTTNMPTDQLEMMVOLCGASVVKELSSFTL 1800
Db 1741 VNGRNHQPKRARSQDRKIFRGLIEICCYGPTTNMPTDQLEMMVOLCGASVVKELSSFTL 1800
QY 1801 GTGVHPIVVQPDAWTEDNGFHAIGQMCAPVVTREWVLDVALYQCQELDTYLIPIPH 1860
Db 1801 GTGVHPIVVQPDAWTEDNGFHAIGQMCAPVVTREWVLDVALYQCQELDTYLIPIPH 1860
QY 1861 SHY 1863
Db 1861 SHY 1863
RESULT 14
US-08-798-691-2
; Sequence 2, Application US/08798691
; Patent No. 5750400
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; APPLICANT: Allen, Antoinette C.
; APPLICANT: Alvares, Christopher P.
; APPLICANT: Critz, Brenda S.
; APPLICANT: Olson, Sheri J.
; APPLICANT: Schelter, Denise B.
; APPLICANT: Zeug, Bin
; TITLE OF INVENTION: Coding Sequences of the Human
; TITLE OF INVENTION: BRC1 Gene
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ONCORMED

STREET: 200 Perry Parkway
CITY: Gaithersburg
STATE: MD
COUNTRY: USA
ZIP: 20877
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/798,691
FILING DATE: 12-Feb-97
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Thomas Gallegos
REGISTRATION NUMBER: 32,692
REFERENCE/DOCKET NUMBER: PA-0054CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-527-2051
TELEFAX: 301-208-6997
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
STRAIN: BRCAL
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 17
MAP POSITION: 17q21

US-08-798-691-2

Query Match 99.9%; Score 9635; DB 1; Length 1863;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1860; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MDLSALRVEEVQNVINAMQKILECPICLELKEPVTCKDHIKFCMCKLLNKKGPQ 60
DB 1 MDLSALRVEEVQNVINAMQKILECPICLELKEPVTCKDHIKFCMCKLLNKKGPQ 60
QY 61 CPLCKNDITKRSLOESTRFSQVVEELKIIICAFQDITGLYANSYNFAKKENNSPEHLKD 120
DB 61 CPLCKNDITKRSLOESTRFSQVVEELKIIICAFQDITGLYANSYNFAKKENNSPEHLKD 120
QY 121 EYSIIQSMGYRNRKRLQSEPNPSLQETSLSVQSLNLTGTVRTLTKRIQPKTSVYI 180
DB 121 EYSIIQSMGYRNRKRLQSEPNPSLQETSLSVQSLNLTGTVRTLTKRIQPKTSVYI 180
QY 181 ELGSDSSEDTVNKATYCSVGDDQLQITPQGTREISLDSAKKAACEFSETDVTNTEHQ 240
DB 181 ELGSDSSEDTVNKATYCSVGDDQLQITPQGTREISLDSAKKAACEFSETDVTNTEHQ 240
QY 241 PSNNDLNTTEKRAAERHPKYGSSVSNLHVPCGNTTHASSLQENSSLLLTKDRMNYE 300
DB 241 PSNNDLNTTEKRAAERHPKYGSSVSNLHVPCGNTTHASSLQENSSLLLTKDRMNYE 300
QY 301 KAEECNKSKOPGLARSQHRNAGSKETCNDRRTPSPTEKKVDLNADPLCKERKNKQKLP 360
DB 301 KAEECNKSKOPGLARSQHRNAGSKETCNDRRTPSPTEKKVDLNADPLCKERKNKQKLP 360
QY 361 SENPRTEDPVITLNSSTQKNEWFSRSDLLGSDSDHGESSENAKVADVDLVINEVD 420
DB 361 SENPRTEDPVITLNSSTQKNEWFSRSDLLGSDSDHGESSENAKVADVDLVINEVD 420
QY 421 EYSGSEKIDLLASDPHEALICKSERVSHKSVESNIEDKIFGKTYRKKASLPNLSHVTEN 480
DB 421 EYSGSEKIDLLASDPHEALICKSERVSHKSVESNIEDKIFGKTYRKKASLPNLSHVTEN 480
QY 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPEMINQCTNOTE 540

DB 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPEMINQCTNOTE 540
QY 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAPPISSISNMLELNI 600
DB 541 QNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAPPISSISNMLELNI 600
QY 601 HNSKAPKNRLRRKSSSTRHIIHALELVVSRNLSPPNCTELQIDSCSSSEIEIKKKYNQMPV 660
DB 601 HNSKAPKNRLRRKSSSTRHIIHALELVVSRNLSPPNCTELQIDSCSSSEIEIKKKYNQMPV 660
QY 661 RHSRNQLMEGKEPATGAKKSNKPNEQTSKRHDSDTFPELKLITNAPGSTKCSNTSELKE 720
DB 661 RHSRNQLMEGKEPATGAKKSNKPNEQTSKRHDSDTFPELKLITNAPGSTKCSNTSELKE 720
QY 721 FVNPSPREEKEKLETVKVSNNADPKDMLSGERVLQTERSVESSTLSLPGDYDGTQ 780
DB 721 FVNPSPREEKEKLETVKVSNNADPKDMLSGERVLQTERSVESSTLSLPGDYDGTQ 780
QY 781 ESISLLEVTGLKAKTEPNKCVSQAAPFNPKGLIHGCSKDNNDTEGFKYPLGHEVNH 840
DB 781 ESISLLEVTGLKAKTEPNKCVSQAAPFNPKGLIHGCSKDNNDTEGFKYPLGHEVNH 840
QY 841 RETSIEMESELDAQYLQNTFKVSRQSFALFSPGNAEEECATFSAHSGSLKKQSPKVT 900
DB 841 RETSIEMESELDAQYLQNTFKVSRQSFALFSPGNAEEECATFSAHSGSLKKQSPKVT 900
QY 901 FECEQKEENQGNESNIKPVQTVNITAGVPVVGOKDKPVDNAKCSIKGGSRCFLSSQFRG 960
DB 901 FECEQKEENQGNESNIKPVQTVNITAGVPVVGOKDKPVDNAKCSIKGGSRCFLSSQFRG 960
QY 961 NETGLITPNKHGLLQNPYRIPPLPIKSPVTKCKKNLLEENFEHSHMSPEREMGNENIP 1020
DB 961 NETGLITPNKHGLLQNPYRIPPLPIKSPVTKCKKNLLEENFEHSHMSPEREMGNENIP 1020
QY 1021 STVSTISRNRIENVPFKEASSNINEVSGSTNEVGSSINEIGSSDENIOAELGRNRGPKL 1080
DB 1021 STVSTISRNRIENVPFKEASSNINEVSGSTNEVGSSINEIGSSDENIOAELGRNRGPKL 1080
QY 1081 NAWLRGLVQPEVYKQSLPGSNCKHPEIKKQEEVYVQVNTDFSPYLLSDNLQPMGSS 1140
DB 1081 NAWLRGLVQPEVYKQSLPGSNCKHPEIKKQEEVYVQVNTDFSPYLLSDNLQPMGSS 1140
QY 1141 HASQVCSETPDDLLDGEIKEDTSAENDIKESSAVFSKQKELSRSPSPFTHTHQAQ 1200
DB 1141 HASQVCSETPDDLLDGEIKEDTSAENDIKESSAVFSKQKELSRSPSPFTHTHQAQ 1200
QY 1201 GYRRGAKKLESSEENLSSDEELPCFQHLFGKVNIPSQSTRHSTVATECLSKNTEENL 1260
DB 1201 GYRRGAKKLESSEENLSSDEELPCFQHLFGKVNIPSQSTRHSTVATECLSKNTEENL 1260
QY 1261 LSLKNSLNDCSNOVILAKASQEHHLSEETKCSASLFSQSCSELEDLTANTNTQDPFLTGS 1320
DB 1261 LSLKNSLNDCSNOVILAKASQEHHLSEETKCSASLFSQSCSELEDLTANTNTQDPFLTGS 1320
QY 1321 SKOMRHQESQGVGLSDKELVSDDEERGTLLEENQEOGSDSNLGEAASGESSTSVSE 1380
DB 1321 SKOMRHQESQGVGLSDKELVSDDEERGTLLEENQEOGSDSNLGEAASGESSTSVSE 1380
QY 1381 DCSGLSSQSDIILTTQORDTMQHNLIKLOQEMAELBAVLQEHQSGQSPSNYSPIISDSSALE 1440
DB 1381 DCSGLSSQSDIILTTQORDTMQHNLIKLOQEMAELBAVLQEHQSGQSPSNYSPIISDSSALE 1440
QY 1441 DLNRNPOSTSEKAVLTQKSSSEYPIQSONPEGLSADKFEVSADSSSTSKNKEPVERSSPSK 1500
DB 1441 DLNRNPOSTSEKAVLTQKSSSEYPIQSONPEGLSADKFEVSADSSSTSKNKEPVERSSPSK 1500
QY 1501 CPSLDDRWMHSCSGSLQNRNYPQOEELIKVVDVPEQOOLEESGPHDLTETSYLPQDLEG 1560
DB 1501 CPSLDDRWMHSCSGSLQNRNYPQOEELIKVVDVPEQOOLEESGPHDLTETSYLPQDLEG 1560
QY 1561 TPYLESGISLFDSDPSPEDRAPESARVGNIPSSTSALKVPOLKVAESAQSAHAHTT 1620

Db 1561 TPYLESGISLFSDDPESDRAPESARVGNIPSTSAALKVPQLKVAESAQQPAAHTT 1620
Qy 1621 DTAGNAMEESVSREKPELTASTERVNKRMSVMVSGLTPEEFMLVYKFAKKHHTLTNLI 1680
Db 1621 DTAGNAMEESVSREKPELTASTERVNKRMSVMVSGLTPEEFMLVYKFAKKHHTLTNLI 1680
Qy 1681 TEETHVVMKTDAAEVCERTLKYFLGIAGGKVVSYFWVTQSIKERKMLNHDPEVRGDV 1740
Db 1681 TEETHVVMKTDAAEVCERTLKYFLGIAGGKVVSYFWVTQSIKERKMLNHDPEVRGDV 1740
Qy 1741 VGNRHQGPKRARESQRKIFRGLIEICCYGPFTNMPTDQLEMMVOLCGASVVKELSSFTL 1800
Db 1741 VGNRHQGPKRARESQRKIFRGLIEICCYGPFTNMPTDQLEMMVOLCGASVVKELSSFTL 1800
Qy 1801 GTGVPIVVVOPDAWTDNGFHAIGOMCEAPVVTREWVLDSVALYQCOELDTYLIPOIPH 1860
Db 1801 GTGVPIVVVOPDAWTDNGFHAIGOMCEAPVVTREWVLDSVALYQCOELDTYLIPOIPH 1860
Qy 1861 SHY 1863
Db 1861 SHY 1863

RESULT 15
US-08-798-691-6
; Sequence 6, Application US/08798691
; Patent No. 5750400
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; APPLICANT: Allen, Antonette C.
; APPLICANT: Alvares, Christopher P.
; APPLICANT: Critz, Brenda S.
; APPLICANT: Olson, Sheri J.
; APPLICANT: Schelter, Denise B.
; APPLICANT: Zeng, Bin
; TITLE OF INVENTION: Coding Sequences of the Human
; TITLE OF INVENTION: BRCAL Gene
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ONCORMED
; STREET: 200 Perry Parkway
; CITY: Gaithersburg
; STATE: MD
; COUNTRY: USA
; ZIP: 20877
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/798,691
; FILING DATE: 12-Feb-97
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Thomas Gallegos
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: PA-0054CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-527-2051
; TELEFAX: 301-208-6997
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCAL
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17

MAP POSITION: 17q21
US-08-798-691-6
Query Match 99.9%; Score 9635; DB 1; Length 1863;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1860; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 MDLSALRVEEVQNVINAMOKILECPICLELTKEPVSTKCDHIFCKFCMLKLLNKKGPSQ 60
Db 1 MDLSALRVEEVQNVINAMOKILECPICLELTKEPVSTKCDHIFCKFCMLKLLNKKGPSQ 60
Qy 61 CPLCKNDITKRSLOESTRFSQLVBEELKIIICAFOLDTGLEVANSYNFAKKENNSPEHLKD 120
Db 61 CPLCKNDITKRSLOESTRFSQLVBEELKIIICAFOLDTGLEVANSYNFAKKENNSPEHLKD 120
Qy 121 EVSIQSMGYRNRAKRLQSEPNPSLOETSLVSOLSNLGTVRTLTKQRTQPKQTSYVI 180
Db 121 EVSIQSMGYRNRAKRLQSEPNPSLOETSLVSOLSNLGTVRTLTKQRTQPKQTSYVI 180
Qy 181 ELGSDSSEDVTNKATYCSVGDOELLQITPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Db 181 ELGSDSSEDVTNKATYCSVGDOELLQITPOGTRDEISLDSAKKAACEFSETDVTNTEHHQ 240
Qy 241 PSNNDLNTTEKRAAERHPEKYQGSVSNLHVPCGTINTHASSLQHENSLLLTDKRMNVE 300
Db 241 PSNNDLNTTEKRAAERHPEKYQGSVSNLHVPCGTINTHASSLQHENSLLLTDKRMNVE 300
Qy 301 KAFCNKSKQPGGLARSQHNHWAGSKETCNDRRTPSTTEKKVVDLNADPLCERKEWNKQKLPC 360
Db 301 KAFCNKSKQPGGLARSQHNHWAGSKETCNDRRTPSTTEKKVVDLNADPLCERKEWNKQKLPC 360
Qy 361 SENPRDTEVPWITLNSIIQVNEWFSRDELIGSDSDSHDGESESNKAVADVLDVNLNVD 420
Db 361 SENPRDTEVPWITLNSIIQVNEWFSRDELIGSDSDSHDGESESNKAVADVLDVNLNVD 420
Qy 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVSVEINIEDKIFGKTYRKASLPNLSHVTE 480
Db 421 EYSGSSEKIDLLASDPHEALICKSERVHKSVSVEINIEDKIFGKTYRKASLPNLSHVTE 480
Qy 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPMINOGTNOTE 540
Db 481 LIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPMINOGTNOTE 540
Qy 541 QNGOVNMTNSGHENKTKGDSIONEKNPNPTIESLEKESAFKTKAEPISSSISNMELELNI 600
Db 541 QNGOVNMTNSGHENKTKGDSIONEKNPNPTIESLEKESAFKTKAEPISSSISNMELELNI 600
Qy 601 HNSKAPKKNLRRKSSRPHIHALELVVSRNLSPPNCTELQIDSCSSSBEIEIKKKYNQMPV 660
Db 601 HNSKAPKKNLRRKSSRPHIHALELVVSRNLSPPNCTELQIDSCSSSBEIEIKKKYNQMPV 660
Qy 661 RHSRNLQMECKEPATGAKKSNKPNEQTSKRHSDSTPELKLTNAPGSFTKCSNTSELKE 720
Db 661 RHSRNLQMECKEPATGAKKSNKPNEQTSKRHSDSTPELKLTNAPGSFTKCSNTSELKE 720
Qy 721 FVNPSPREEKEEKLETVKVSNNNAEDPKDMLSGERVLTQTSRVSSESSISLVPGTDYGTQ 780
Db 721 FVNPSPREEKEEKLETVKVSNNNAEDPKDMLSGERVLTQTSRVSSESSISLVPGTDYGTQ 780
Qy 781 ESISLLEVTGLGAKTEPNKCVSOCAAFENPKGLIHGCSKDNDRNDTEGFKYPLGHEVNHS 840
Db 781 ESISLLEVTGLGAKTEPNKCVSOCAAFENPKGLIHGCSKDNDRNDTEGFKYPLGHEVNHS 840
Qy 841 RETSIEMESELDAQYLQNTFKVSKRQSFALFSPNPGNAEBECATFSAHSGSLKKQSPKVT 900
Db 841 RETSIEMESELDAQYLQNTFKVSKRQSFALFSPNPGNAEBECATFSAHSGSLKKQSPKVT 900
Qy 901 FECEQKEENOGKNESNIKPVQTNITAGFPVVGOKDPVDNAKCSIKGGSFCLSSQFRG 960
Db 901 FECEQKEENOGKNESNIKPVQTNITAGFPVVGOKDPVDNAKCSIKGGSFCLSSQFRG 960
Qy 961 NETGLITPNKHGLLQNPYRIPPLFIKSVYTKCKKNLLEENFEHSMSPREMGNENIP 1020
Db 961 NETGLITPNKHGLLQNPYRIPPLFIKSVYTKCKKNLLEENFEHSMSPREMGNENIP 1020

Db	961	NETGLITPNKHGLLQNLQNYRIPPLPIKSFVKTKCKKNLLENFEEHSMSPEREMGNENIP	1020
Qy	1021	STVSTISRNIRENVFKEASSNINEVGSSTNEIGSSINIEIGSSDENITQAEIAGRNRPKL	1080
Db	1021	STVSTISRNIRENVFKEASSNINEVGSSTNEIGSSINIEIGSSDENITQAEIAGRNRPKL	1080
Qy	1081	NAMLRLGVLOPEVYKQSLPGSNCKHPKIKQOEYEVVQTVNTDFSPYLIISDNLEQPMGSS	1140
Db	1081	NAMLRLGVLOPEVYKQSLPGSNCKHPKIKQOEYEVVQTVNTDFSPYLIISDNLEQPMGSS	1140
Qy	1141	HASQVCSETPDDLDDGCEIREDTSAFENDIKESSAVFSKSVOKGELSRSPFTHHLAQ	1200
Db	1141	HASQVCSETPDDLDDGCEIREDTSAFENDIKESSAVFSKSVOKGELSRSPFTHHLAQ	1200
Qy	1201	GYRGAKKLESSEENLSSEDEELPCFOHLLFGKVNNIPQSOTRHSSTVATECLSKNTEENL	1260
Db	1201	GYRGAKKLESSEENLSSEDEELPCFOHLLFGKVNNIPQSOTRHSSTVATECLSKNTEENL	1260
Qy	1261	LSLKNLSNDCSNQVILAKASQEHHLSEETKCSASLFSQCSQSELEDLTANTNTQDPFLIGS	1320
Db	1261	LSLKNLSNDCSNQVILAKASQEHHLSEETKCSASLFSQCSQSELEDLTANTNTQDPFLIGS	1320
Qy	1321	SKOMRHOSQGVGLSDKELVSDDEERTGTGLENNQEQSMDSNLGEAASCCSETSVSE	1380
Db	1321	SKOMRHOSQGVGLSDKELVSDDEERTGTGLENNQEQSMDSNLGEAASCCSETSVSE	1380
Qy	1381	DCSGLSSQSDILTTQQRDTMQHNLIKIQOEMAELEAVLEQHGQSPSNSYPSIISDSSALE	1440
Db	1381	DCSGLSSQSDILTTQQRDTMQHNLIKIQOEMAELEAVLEQHGQSPSNSYPSIISDSSALE	1440
Qy	1441	DLRNPQOSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSTSKNKEPGEVERSSPSK	1500
Db	1441	DLRNPQOSTSEKAVLTSQKSSEYPISONPEGLSADKFEVSADSTSKNKEPGEVERSSPSK	1500
Qy	1501	CPSLDDRWMYHSCSGSLQNRNYPQOEELIKVVDVEEQOLEESGPHDLTETSYLPRQDLEG	1560
Db	1501	CPSLDDRWMYHSCSGSLQNRNYPQOEELIKVVDVEEQOLEESGPHDLTETSYLPRQDLEG	1560
Qy	1561	TPYLESIGISLFSDDPESDRAPE SARVGNIPSTTSALKVPOLKVAESAQSPAAAHTT	1620
Db	1561	TPYLESIGISLFSDDPESDRAPE SARVGNIPSTTSALKVPOLKVAESAQSPAAAHTT	1620
Qy	1621	DTAGYNAMESVSREKPELFASTERNVKNMSVVSGLTPEEFMLVYKFAKKHHTLTNLI	1680
Db	1621	DTAGYNAMESVSREKPELFASTERNVKNMSVVSGLTPEEFMLVYKFAKKHHTLTNLI	1680
Qy	1681	TEETHVVMKTDAEFVCERTIKYFLGIAGGKWVVS YFWVTQSIKERKMLNEHDFEVRGDV	1740
Db	1681	TEETHVVMKTDAEFVCERTIKYFLGIAGGKWVVS YFWVTQSIKERKMLNEHDFEVRGDV	1740
Qy	1741	VNGRNHOGPKRARESQDRKIFRGLIEICCYGFTNMPTDQLEWMVOLCGASVVKELSSFTL	1800
Db	1741	VNGRNHOGPKRARESQDRKIFRGLIEICCYGFTNMPTDQLEWMVOLCGASVVKELSSFTL	1800
Qy	1801	GTGVHPITVVVQPDWEDNGFHAIGOMCEAPVVTREWVLDSVALYQCOQELDTYLIPQIPH	1860
Db	1801	GTGVHPITVVVQPDWEDNGFHAIGOMCEAPVVTREWVLDSVALYQCOQELDTYLIPQIPH	1860
Qy	1861	SHY 1863	
Db	1861	SHY 1863	

